

On all things star-free

Thomas Place

LaBRI, Bordeaux University and IUF, France

Marc Zeitoun

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France

Abstract

We investigate the star-free closure, which associates to a class of languages its closure under Boolean operations and marked concatenation. We prove that the star-free closure of any finite class and of any class of groups languages with decidable separation (plus mild additional properties) has decidable separation. We actually show decidability of a stronger property, called covering. This generalizes many results on the subject in a unified framework. A key ingredient is that star-free closure coincides with another closure operator where Kleene stars are also allowed in restricted contexts.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory

Keywords and phrases Regular languages, separation problem, star-free closure, group languages

Digital Object Identifier 10.4230/LIPIcs.ICALP.2019.121

Funding DeLTA project (ANR-16-CE40-0007)

1 Introduction

This paper investigates a remarkable operation on classes of languages: the *star-free closure*. It builds a new class $SF(\mathcal{C})$ from an input class \mathcal{C} by closing it under union, complement and concatenation. This generalizes an important specific class: the one of *star-free languages*, *i.e.*, the star-free closure of the class consisting of all finite languages. Star-free languages are those that can be defined in first order logic [12]. The correspondence was lifted to the quantifier alternation hierarchy of first order logic by Thomas [30], which corresponds to a classification of star-free languages: the dot-depth hierarchy [4]. These results extend to the star-free closure [23]. For each input class \mathcal{C} , $SF(\mathcal{C})$ corresponds to a variant of first-order logic (specified by the set of predicates that are allowed). Moreover, its quantifier alternation hierarchy corresponds to a classification of $SF(\mathcal{C})$: the concatenation hierarchy of basis \mathcal{C} .

Schützenberger proved that one may decide whether a regular language is star-free [27]. This result established a framework for investigating and understanding classes of languages, based on the *membership problem*: is it decidable to test whether an input regular language belongs to the class under investigation? Similar results were obtained for other prominent classes. Yet, this fruitful line of research also includes some of the most famous open problems in automata theory. For example, only the first levels of the dot-depth hierarchy are known to have decidable membership (see [14] for a survey).

Recently, these results were unified and generalized. First, the problem itself was strengthened: membership was replaced by separation as a means to investigate classes. The separation problem asks whether two input languages can be separated by one from the class under study. While more general and difficult than membership, separation is also more flexible. This was exploited to show that separation is decidable for several levels in the dot-depth hierarchy [19, 17]. In fact, this is a particular instance of a *generic* result applying to every hierarchy whose basis \mathcal{C} is *finite* and satisfies some mild properties [18, 21]. Moreover, the same result was obtained when the basis \mathcal{C} is a class of group languages (*i.e.*, recognized by a finite group) with decidable separation [26]. Altogether, these results generalize most of the known results regarding the decidability of levels in concatenation hierarchies.



© Thomas Place and Marc Zeitoun;

licensed under Creative Commons License CC-BY

46th International Colloquium on Automata, Languages, and Programming (ICALP 2019).

Editors: Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi;

Article No. 121; pp. 121:1–121:39



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



Contributions. This paper is a continuation of these research efforts. Instead of looking at levels within hierarchies, we investigate the star-free closure as a whole. First, we show that the star-free closure of a finite class has decidable separation. We then use this result to establish our main theorem: the star-free closure of a class of group languages with *decidable separation* has also decidable separation. In both cases, we actually prove the decidability of a stronger property called covering. Let us mention some important features of this work.

A first point is that the case of a finite class is important by itself. Foremost, it is a crucial step for the main result on the star-free closure of classes of group languages. Second, it yields a new proof that covering is decidable for the star-free languages (this is shown in [20] or can be derived from [9, 1]). This new proof is simpler and generic. While the original underlying technique goes back to Wilke [31], the proof has been simplified at several levels. The main simplification is obtained thanks to an abstract framework, introduced in [22]. It is based on the central notion of rating map, which is meant to measure the quality of a separator. For the framework to be relevant, we actually need to generalize separation to multiple input languages, which leads to the covering problem. Another key difference is that previously existing proofs (specific to the star-free languages) involve abstracting words by new letters at some point, which requires a relabeling procedure and a change of alphabet. Here, we cannot use this approach as the classes we build with star-free closure are less robust in general. We work with a fixed alphabet, which also makes the proof simpler.

A crucial ingredient in the proof is the notion of prefix code with bounded synchronization delay. Generalizing a definition of Schützenberger [28] which was also considered by Diekert and Walter [6, 7], we define a new closure operator that permits Kleene stars on such languages (this is a semantic property). This yields an operator that happens to coincide with the star-free closure when applied to the classes that we investigate. It serves as a key intermediary: in our proofs, we heavily rely on Kleene stars to construct languages. We therefore present this important step in the body of the paper (Theorem 7). Moreover, its proof provides yet another characterization of $SF(\mathcal{C})$, which is effective when the class \mathcal{C} is finite (thus generalizing Schützenberger’s membership result). At last regarding membership, it is worth pointing out that not only do we cover more cases, but also that it is straightforward to reprove the known algebraic characterizations from our results (see *e.g.*, [3]).

Finally, let us present important applications of our main result applying to input classes made of group languages. First, one may look at the input class containing all group languages. Straubing [29] described an algebraic counterpart of the star-free closure of this class, which was then shown to be recursive by Rhodes and Karnofsky [10]. Altogether, this implies that membership is decidable for the star-free closure of group languages, as noted by Margolis and Pin [11]. Here, we are able to generalize this result to separation and covering as separation is known to be decidable for the group languages [2].

Another important application is the class of languages definable by first-order logic with modular predicates $FO(<, \text{MOD})$. This class is known to have decidable membership [3]. Moreover, it is the star-free closure of the class consisting of the languages counting the length of words modulo some number. Since this input class is easily shown to have decidable separation (see [26] for example), our main theorem applies.

The third application applies to first-order logic endowed with predicates counting the number of occurrences of a letter before a position modulo some integer. The languages definable in this logic form the star-free closure of the languages recognized by Abelian groups: this follows from a generic correspondence between star-free closure and variants of first-order logic [23, 13], together with the description of languages recognized by Abelian groups [8]. Our main theorem applies, since the class of Abelian groups has decidable separation [5, 1].

Organization. In Section 2, we recall some useful background. Section 3 presents a generic characterization of star-free closure. Then, Sections 4 and 5 are devoted to our two main theorems applying respectively to finite input classes and those made of group languages. Due to space limitations, several proofs are postponed to the full version of the paper [25].

2 Preliminaries

We fix a finite alphabet A for the whole paper. As usual, A^* denotes the set of all words over A , including the empty word ε . For $u, v \in A^*$, we denote by uv the word obtained by concatenating u and v . A *language* is a subset of A^* . We lift concatenation to languages: for $K, L \subseteq A^*$, we let $KL = \{uv \mid u \in K \text{ and } v \in L\}$. Finally, we use Kleene star: if $K \subseteq A^*$, K^+ denotes the union of all languages K^n for $n \geq 1$ and $K^* = K^+ \cup \{\varepsilon\}$.

A *class of languages* is a set of languages. A class \mathcal{C} is a *Boolean algebra* when it is closed under union, intersection and complement. Moreover, \mathcal{C} is *quotient-closed* if for every $L \in \mathcal{C}$ and $w \in A^*$, the languages $w^{-1}L \stackrel{\text{def}}{=} \{u \in A^* \mid wu \in L\}$ and $Lw^{-1} \stackrel{\text{def}}{=} \{u \in A^* \mid uw \in L\}$ belong to \mathcal{C} . All classes considered in the paper are quotient-closed Boolean algebras containing only *regular languages* (this will be implicit in our statements). These are the languages that can be equivalently defined by monadic second-order logic, finite automata or finite monoids. We briefly recall the monoid-based definition below.

We shall often consider *finite* quotient-closed Boolean algebras. If \mathcal{C} is such a class, one may associate a canonical equivalence $\sim_{\mathcal{C}}$ over A^* . For $w, w' \in A^*$, $w \sim_{\mathcal{C}} w'$ if and only if $w \in L \Leftrightarrow w' \in L$ for every $L \in \mathcal{C}$. Moreover, we write $[w]_{\mathcal{C}} \in A^*/\sim_{\mathcal{C}}$ for the $\sim_{\mathcal{C}}$ -class of w . One may then verify that the languages in \mathcal{C} are exactly the unions of $\sim_{\mathcal{C}}$ -classes. Moreover, since \mathcal{C} is quotient-closed, $\sim_{\mathcal{C}}$ is a congruence for word concatenation (see [23] for proofs).

Regular languages. A *monoid* is a set M endowed with an associative multiplication $(s, t) \mapsto s \cdot t$ (also denoted by st) having a neutral element 1_M . An *idempotent* of a monoid M is an element $e \in M$ such that $ee = e$. It is folklore that for any *finite* monoid M , there exists a natural number $\omega(M)$ (denoted by ω when M is understood) such that s^ω is an idempotent for every $s \in M$. Observe that A^* is a monoid whose multiplication is concatenation (the neutral element is ε). Thus, we may consider monoid morphisms $\alpha : A^* \rightarrow M$ where M is an arbitrary monoid. Given such a morphism and $L \subseteq A^*$, we say that L is *recognized* by α when there exists a set $F \subseteq M$ such that $L = \alpha^{-1}(F)$. A language L is *regular* if and only if it is recognized by a morphism into a *finite* monoid. Moreover, it is known that there exists a canonical recognizer of L , which can be computed from any representation of L (such as a finite automaton): the syntactic morphism of L . We refer the reader to [15] for details.

Group languages. A group is a monoid G in which every element $g \in G$ has an inverse $g^{-1} \in G$, i.e., $gg^{-1} = g^{-1}g = 1_G$. A “*group language*” is a language L recognized by a morphism into a *finite group*. All classes of group languages investigated here are quotient-closed Boolean algebras. Typically, publications on the topic consider *varieties* of group languages which is more restrictive: they involve an additional closure property called “inverse morphic image” (see [13]). For example, the class MOD described below is *not* a variety.

► **Example 1.** A simple example of quotient-closed Boolean algebra of group languages is the class of *all* group languages: GR. Another one is MOD, which contains the Boolean combinations of languages $\{w \in A^* \mid |w| = k \pmod m\}$ with $k, m \in \mathbb{N}$ such that $k < m$.

Decision problems. We rely on three decision problems to investigate classes of languages. Each one depends on a parameter class \mathcal{C} , which we fix for the definition. The first problem, *\mathcal{C} -membership*, takes a single regular language L as input and asks whether $L \in \mathcal{C}$.

The second one, \mathcal{C} -separation, takes two regular languages L_1 and L_2 as input and asks whether L_1 is \mathcal{C} -separable from L_2 (is there a third language $K \in \mathcal{C}$ such that $L_1 \subseteq K$ and $L_2 \cap K = \emptyset$). This generalizes membership: $L \in \mathcal{C}$ if and only if L is \mathcal{C} -separable from $A^* \setminus L$.

The third problem, \mathcal{C} -covering was introduced in [22]. Given a language L , a *cover* of L is a **finite** set of languages \mathbf{K} such that $L \subseteq \bigcup_{K \in \mathbf{K}} K$. Moreover, a \mathcal{C} -cover of L is a cover \mathbf{K} of L such that all $K \in \mathbf{K}$ belong to \mathcal{C} . Consider a pair (L_1, \mathbf{L}_2) where L_1 is a language and \mathbf{L}_2 is a *finite set of languages*. We say that (L_1, \mathbf{L}_2) is \mathcal{C} -coverable when there exists a \mathcal{C} -cover \mathbf{K} of L_1 such that for every $K \in \mathbf{K}$, there exists $L \in \mathbf{L}_2$ satisfying $K \cap L = \emptyset$. The \mathcal{C} -covering problem takes as input a single regular language L_1 and a finite set of regular languages \mathbf{L}_2 . It asks whether (L_1, \mathbf{L}_2) is \mathcal{C} -coverable. Covering generalizes separation if \mathcal{C} is closed under union: L_1 is \mathcal{C} -separable from L_2 , if and only if $(L_1, \{L_2\})$ is \mathcal{C} -coverable (see [22]).

Star-free closure and main results. We investigate an operation defined on classes: *star-free closure*. Consider a class \mathcal{C} . The *star-free closure* of \mathcal{C} , denoted by $SF(\mathcal{C})$, is the least class containing \mathcal{C} and the singletons $\{a\}$ for every $a \in A$, and closed under Boolean operations and concatenation. It is standard and simple to verify that when \mathcal{C} is a quotient-closed Boolean algebra (which will always be the case here), this is also the case for $SF(\mathcal{C})$.

Our main theorems state conditions on the input class \mathcal{C} guaranteeing decidability of our decision problems for $SF(\mathcal{C})$. First, we may handle *finite* classes.

► **Theorem 2.** *Let \mathcal{C} be a finite quotient-closed Boolean algebra. Then, membership, separation and covering are decidable for $SF(\mathcal{C})$.*

The second theorem applies to input classes made of group languages.

► **Theorem 3.** *Let \mathcal{C} be a quotient-closed Boolean algebra of group languages with decidable separation. Then, membership, separation and covering are decidable for $SF(\mathcal{C})$.*

The remainder of the paper is devoted to proving these theorems. We first focus on $SF(\mathcal{C})$ -membership in Section 3. Naturally, this is weaker than directly handling $SF(\mathcal{C})$ -covering. Yet, detailing membership independently allows to introduce many proof ideas and techniques that are needed to prove the “full” theorems. We detail these theorems in Sections 4 and 5. We only present the algorithms: proofs are deferred to the full paper [25].

3 Bounded synchronization delay and algebraic characterization

This section is devoted to $SF(\mathcal{C})$ -membership. We handle it with a generic algebraic characterization of $SF(\mathcal{C})$ (effective under the hypotheses of Theorems 2 and 3), generalizing earlier work by Pin, Straubing and Thérien [29, 16]. We rely on an alternate definition of star-free closure involving a semantic restriction of the Kleene star, which we first present.

3.1 Bounded synchronization delay

We define a second operation on classes of languages $\mathcal{C} \mapsto SD(\mathcal{C})$. We shall later prove that it coincides with star-free closure (provided that \mathcal{C} satisfies mild hypotheses). It is based on the work of Schützenberger [28] who defined a single class SD corresponding to the star-free languages (*i.e.*, $SF(\{\emptyset, A^*\})$). Here, we generalize it as an operation. The definition involves a semantic restriction of the Kleene star operation on languages: it may only be applied to “*prefix codes with bounded synchronization delay*”. Introducing this notion requires basic definitions from coding theory that we first recall.

A language $K \subseteq A^*$ is a *prefix code* when $\varepsilon \notin K$ and $K \cap KA^+ = \emptyset$ (no word in K has a strict prefix in K). Note that this implies the following weaker property that we shall use implicitly: every $w \in K^*$ admits a *unique* decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in K$ (this property actually defines *codes* which are more general).

Given $d \geq 1$, a prefix code $K \subseteq A^+$ has *synchronization delay* d if for every $u, v, w \in A^*$ such that $uvw \in K^+$ and $v \in K^d$, we have $uv \in K^+$. Finally, a prefix code $K \subseteq A^+$ has *bounded synchronization delay* when it has synchronization delay d for some $d \geq 1$.

► **Example 4.** Let $A = \{a, b\}$. Clearly, $\{ab\}$ is a prefix code with synchronization delay 1: if $uvw \in (ab)^+$ and $v = ab$, we have $uv \in (ab)^+$. Similarly, one may verify that $(aab)^*ab$ is a prefix code with synchronization delay 2 (but not 1). On the other hand, $\{aa\}$ does not have bounded synchronization delay. If $d \geq 1$, $a(aa)^d a \in (aa)^*$ but $a(aa)^d \notin (aa)^*$.

We present the operation $\mathcal{C} \mapsto SD(\mathcal{C})$. The definition involves *unambiguous concatenation*. Given $K, L \subseteq A^*$, their concatenation KL is *unambiguous* when every word $w \in KL$ admits a *unique* decomposition $w = uv$ with $u \in K$ and $v \in L$. Given a class \mathcal{C} , $SD(\mathcal{C})$ is the least class containing \emptyset and $\{a\}$ for every $a \in A$, and closed under the following properties:

- **Intersection with \mathcal{C} :** if $K \in SD(\mathcal{C})$ and $L \in \mathcal{C}$, then $K \cap L \in SD(\mathcal{C})$.
- **Disjoint union:** if $K, L \in SD(\mathcal{C})$ are disjoint, then $K \uplus L \in SD(\mathcal{C})$.
- **Unambiguous product:** if $K, L \in SD(\mathcal{C})$ and KL is unambiguous, then $KL \in SD(\mathcal{C})$.
- **Kleene star for prefix codes with bounded synchronization delay:** if $K \in SD(\mathcal{C})$ is a prefix code with bounded synchronization delay, then $K^* \in SD(\mathcal{C})$.

► **Remark 5.** Schützenberger proved in [28] that $SD(\{\emptyset, A^*\}) = SF(\{\emptyset, A^*\})$. His definition of $SD(\{\emptyset, A^*\})$ was slightly less restrictive than ours: it does not require that the unions are disjoint and the concatenations unambiguous. It will be immediate from the correspondence with star-free closure that the two definitions are equivalent.

► **Remark 6.** This closure operation is different from standard ones. Instead of requiring that $\mathcal{C} \subseteq SD(\mathcal{C})$, we impose a stronger requirement: intersection with languages in \mathcal{C} is allowed. If we only asked that $\mathcal{C} \subseteq SD(\mathcal{C})$, we would get a weaker operation which does not correspond to star-free closure in general. For example, let $A = \{a, b\}$ and consider the class MOD of Example 1. Observe that $(aa)^* \in SD(\text{MOD})$. Indeed, $\{a\} \in SD(\text{MOD})$ has bounded synchronization delay, $(AA)^* \in \text{MOD}$ and $(aa)^* = a^* \cap (AA)^*$. Yet, one may verify that $(aa)^*$ cannot be built from the languages of MOD with union, concatenation and Kleene star applied to prefix codes with bounded synchronization delay.

3.2 Algebraic characterization of star-free closure

We now reduce deciding membership for $SF(\mathcal{C})$ to computing \mathcal{C} -stutters. Let us first define this new notion. Let \mathcal{C} be a quotient-closed Boolean algebra and $\alpha : A^* \rightarrow M$ be a morphism. A \mathcal{C} -stutter for α is an element $s \in M$ such that for every \mathcal{C} -cover \mathbf{K} of $\alpha^{-1}(s)$, there exists $K \in \mathbf{K}$ satisfying $K \cap KK \neq \emptyset$. When α is understood, we simply speak of a \mathcal{C} -stutter. Finally, we say that α is \mathcal{C} -aperiodic when for every \mathcal{C} -stutter $s \in M$, we have $s^\omega = s^{\omega+1}$. The reduction is stated in the following theorem.

► **Theorem 7.** *Let \mathcal{C} be a quotient-closed Boolean algebra and consider a regular language $L \subseteq A^*$. The following properties are equivalent:*

1. $L \in SF(\mathcal{C})$.
2. $L \in SD(\mathcal{C})$.
3. *The syntactic morphism of L is \mathcal{C} -aperiodic.*

Naturally, the characterization need not be effective: this depends on \mathcal{C} . Deciding whether a morphism is \mathcal{C} -aperiodic boils down to computing \mathcal{C} -stutters. Yet, this is possible under the hypotheses of Theorems 2 and 3. First, if \mathcal{C} is a finite quotient-closed Boolean algebra, deciding whether an element is a \mathcal{C} -stutter is simple: there are finitely many \mathcal{C} -covers and we may check them all. If \mathcal{C} is a quotient-closed Boolean algebra of group languages, the question boils down to \mathcal{C} -separation as stated in the next lemma (proved in [25]).

► **Lemma 8.** *Let \mathcal{C} be a quotient-closed Boolean algebra of group languages and $\alpha : A^* \rightarrow M$ be a morphism. For all $s \in M$, s is a \mathcal{C} -stutter if and only if $\{\varepsilon\}$ is **not** \mathcal{C} -separable from $\alpha^{-1}(s)$.*

Altogether, we obtain the membership part in Theorems 20 and 25. We conclude the section with an extended proof sketch for the most interesting direction in Theorem 7: $3) \Rightarrow 2)$ (a detailed proof for the two other directions is provided in the full version of this paper [25]).

Proof of $3) \Rightarrow 2)$ in Theorem 7. Let \mathcal{C} be a quotient-closed Boolean algebra and $\alpha : A^* \rightarrow M$ be a \mathcal{C} -aperiodic morphism. We show that all languages recognized by α belong to $SD(\mathcal{C})$.

Given $K \subseteq A^*$ and $s \in M$, we say that K is s -safe when $s\alpha(u) = s\alpha(v)$ for every $u, v \in K$. We extend this notion to sets of languages: such a set \mathbf{K} is s -safe when every $K \in \mathbf{K}$ is s -safe. We shall use s as an induction parameter. Finally, given a language $P \subseteq A^*$, an $SD(\mathcal{C})$ -partition of P is a finite partition of P into languages of $SD(\mathcal{C})$.

► **Proposition 9.** *Let $P \subseteq A^+$ be a prefix code with bounded synchronization delay. Assume that there exists a 1_M -safe $SD(\mathcal{C})$ -partition of P . Then, for every $s \in M$, there exists an s -safe $SD(\mathcal{C})$ -partition of P^* .*

We first apply Proposition 9 to conclude the main argument. We show that every language recognized by α belongs to $SD(\mathcal{C})$. By definition, $SD(\mathcal{C})$ is closed under disjoint union. Hence, it suffices to show that $\alpha^{-1}(t) \in SD(\mathcal{C})$ for every $t \in M$. We fix $t \in M$ for the proof.

Clearly, $A \subseteq A^+$ is a prefix code with bounded synchronization delay and $\{\{a\} \mid a \in A\}$ is a 1_M -safe $SD(\mathcal{C})$ -partition of A . Hence, Proposition 9 (applied in the case $s = 1_M$) yields a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{K} of A^* . One may verify that $\alpha^{-1}(t)$ is the disjoint union of all $K \in \mathbf{K}$ intersecting $\alpha^{-1}(t)$. Hence, $\alpha^{-1}(t) \in SD(\mathcal{C})$ which concludes the main argument.

It remains to prove Proposition 9. We let $P \subseteq A^*$ be a prefix code with bounded synchronization delay, \mathbf{H} a 1_M -safe $SD(\mathcal{C})$ -partition of P and $s \in M$. We need to build an $SD(\mathcal{C})$ -partition \mathbf{K} of P^* such that every $K \in \mathbf{K}$ is s -safe. We proceed by induction on the three following parameters listed by order of importance: (1) the size of $\alpha(P^+) \subseteq M$, (2) the size of \mathbf{H} and (3) the size of $s \cdot \alpha(P^*) \subseteq M$. We distinguish two cases depending on the following property of s and \mathbf{H} . We say that s is **H-stable** when the following holds:

$$\text{for every } H \in \mathbf{H}, \quad s \cdot \alpha(P^*) = s \cdot \alpha(P^*H). \quad (1)$$

The base case happens when s is **H-stable**. Otherwise, we use induction on our parameters.

Base case: s is **H-stable.** Since α is \mathcal{C} -aperiodic, we have the following simple fact.

► **Fact 10.** *There is a finite quotient-closed Boolean algebra $\mathcal{D} \subseteq \mathcal{C}$ such that α is \mathcal{D} -aperiodic.*

Since \mathcal{D} is finite, we may consider the associated canonical equivalence $\sim_{\mathcal{D}}$ over A^* . We let $\mathbf{K} = \{P^* \cap D \mid D \in A^*/\sim_{\mathcal{D}}\}$. Clearly, \mathbf{K} is a partition of P^* . Let us verify that it only contains languages in $SD(\mathcal{C})$. We have $P \in SD(\mathcal{C})$: it is the disjoint union of all languages in the $SD(\mathcal{C})$ -partition \mathbf{H} of P . Moreover, $P^* \in SD(\mathcal{C})$ since P is a prefix code with bounded synchronization delay. Hence, $P^* \cap D \in SD(\mathcal{C})$ for every $D \in A^*/\sim_{\mathcal{D}}$ since $D \in \mathcal{D} \subseteq \mathcal{C}$. Therefore, it remains to show that every language $K \in \mathbf{K}$ is s -safe. This is a consequence of the following lemma which is proved using the hypothesis (1) that s is **H-stable**.

► **Lemma 11.** *For every $u, v \in P^*$ such that $u \sim_{\mathcal{D}} v$, we have $s\alpha(u) = s\alpha(v)$.*

Inductive step: s is not \mathbf{H} -stable. By hypothesis, we know that (1) does not hold. Therefore, we get some $H \in \mathbf{H}$ such that the following **strict** inclusion holds,

$$s \cdot \alpha(P^*H) \subsetneq s \cdot \alpha(P^*). \quad (2)$$

We fix this language $H \in \mathbf{H}$ for the remainder of the proof. The following lemma is proved by induction on our second parameter (the size of \mathbf{H}).

► **Lemma 12.** *There exists a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{U} of $(P \setminus H)^*$.*

We fix the partition \mathbf{U} of $(P \setminus H)^*$ given by Lemma 12 and distinguish two independent sub-cases. Since $H \subseteq P$ (as H is an element of the partition \mathbf{H} of P), we have $\alpha(P^*H) \subseteq \alpha(P^+)$. We use a different argument depending on whether this inclusion is strict or not.

Sub-case 1: $\alpha(P^*H) = \alpha(P^+)$. Since H is 1_M -safe by hypothesis, there exists $t \in M$ such that $\alpha(H) = \{t\}$. Similarly, since every $U \in \mathbf{U}$ is 1_M -safe, there exists $r_U \in M$ such that $\alpha(U) = \{r_U\}$. The construction of \mathbf{K} is based on the next lemma which is proved using (2), the hypothesis of Sub-case 1 and induction on our third parameter (the size of $s \cdot \alpha(P^*) \subseteq M$).

► **Lemma 13.** *For every $U \in \mathbf{U}$, there exists an $sr_U t$ -safe $SD(\mathcal{C})$ -partition \mathbf{W}_U of P^* .*

We are ready to define the partition \mathbf{K} of P^* . Using Lemma 13, we define,

$$\mathbf{K} = \mathbf{U} \cup \bigcup_{U \in \mathbf{U}} \{UHW \mid W \in \mathbf{W}_U\}$$

It remains to show that \mathbf{K} is an s -safe $SD(\mathcal{C})$ -partition of P^* . First, \mathbf{K} is a partition of P^* since P is a prefix code and $H \subseteq P$. Indeed, every word $w \in P^*$ admits a *unique* decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in P$. If no factor w_i belongs to H , then $w \in (P \setminus H)^*$ and w belongs to some unique $U \in \mathbf{U}$. Otherwise, let w_i be the leftmost factor such that $w_i \in H$. Thus, $w_1 \cdots w_{i-1} \in (P \setminus H)^*$, which also yields a unique $U \in \mathbf{U}$ such that $w_1 \cdots w_{i-1} \in U$ and $w_{i+1} \cdots w_n \in P^*$ which yields a unique $W \in \mathbf{W}_U$ such that $w_{i+1} \cdots w_n \in W$. Thus, $w \in UHW$ which is an element of \mathbf{K} (the only one containing w).

Moreover, every $K \in \mathbf{K}$ belongs to $SD(\mathcal{C})$. If $K \in \mathbf{U}$, this is immediate by definition of \mathbf{U} in Lemma 12. Otherwise, $K = UHW$ with $U \in \mathbf{U}$ and $W \in \mathbf{W}_U$. We know that $U, H, W \in SD(\mathcal{C})$ by definition. Moreover, one may verify that the concatenation UHW is *unambiguous* since P is a prefix code, $U \subseteq (P \setminus H)^*$ and $W \subseteq H^*$. Hence, $K \in SD(\mathcal{C})$.

Finally, we verify that \mathbf{K} is s -safe. Consider $K \in \mathbf{K}$ and $w, w' \in K$, we show that $s\alpha(w) = s\alpha(w')$. If $K \in \mathbf{U}$, this is immediate: \mathbf{U} is 1_M -safe by definition. Otherwise, $K = UHW$ with $U \in \mathbf{U}$ and $W \in \mathbf{W}_U$. By definition, $\alpha(H) = \{t\}$ and $\alpha(U) = \{r_U\}$ which implies that $s\alpha(w) = sr_U t\alpha(x)$ and $s\alpha(w') = sr_U t\alpha(x')$ for $x, x' \in W$. Moreover, $W \in \mathbf{W}_U$ is $sr_U t$ -safe by definition. Hence, $s\alpha(w) = s\alpha(w')$, which concludes the proof of this sub-case.

Sub-case 2: $\alpha(P^*H) \subsetneq \alpha(P^+)$. Consider $w \in P^*$. Since P is a prefix code, w admits a unique decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in P$. We may look at the rightmost factor $w_i \in H \subseteq P$ to uniquely decompose w in two parts (each of them possibly empty): the prefix $w_1 \cdots w_i \in ((P \setminus H)^*H)^*$ and the suffix in $w_{i+1} \cdots w_n \in (P \setminus H)^*$. Using induction, we construct $SD(\mathcal{C})$ -partitions of the possible languages of prefixes and suffixes. Then, we combine them to construct a partition of the whole set P^* . We already handled the suffixes: \mathbf{U} is an $SD(\mathcal{C})$ -partition of $(P \setminus H)^*$. The prefixes are handled using the hypothesis of Sub-case 2 and induction on our first parameter (the size of $\alpha(P^+)$).

► **Lemma 14.** *There exists a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{V} of $((P \setminus H)^*H)^*$.*

Using Lemma 14, we define $\mathbf{K} = \{VU \mid V \in \mathbf{V} \text{ and } U \in \mathbf{U}\}$. It follows from the above discussion that \mathbf{K} is a partition of P^* since \mathbf{V} and \mathbf{U} are partitions of $((P \setminus H)^*H)^*$ and $(P \setminus H)^*$, respectively. Moreover, every $K \in \mathbf{K}$ belongs to $SD(\mathcal{C})$: $K = VU$ with $V \in \mathbf{V}$ and $U \in \mathbf{U}$, and one may verify that this is an *unambiguous* concatenation. It remains to show that \mathbf{K} is s -safe. Let $K \in \mathbf{K}$ and $w, w' \in K$. We show that $s\alpha(w) = s\alpha(w')$. By definition, we have $K = VU$ with $V \in \mathbf{V}$ and $U \in \mathbf{U}$. Therefore, $w = vu$ and $w' = v'u'$ with $u, u' \in U$ and $v, v' \in V$. Since U and V are both 1_M -safe by definition, we have $\alpha(u) = \alpha(u')$ and $\alpha(v) = \alpha(v')$. It follows that $s\alpha(w) = s\alpha(w')$, which concludes the proof of Proposition 9. ◀

4 Covering when the input class is finite

This section is devoted to Theorem 2. We show that when \mathcal{C} is a finite quotient-closed Boolean algebra, $SF(\mathcal{C})$ -covering is decidable by presenting a generic algorithm. It is formulated within a framework designed to handle covering questions, which was originally introduced in [22]. We start by briefly recalling it (we refer the reader to [22] for details).

4.1 Rating maps and optimal imprints

The framework is based on an algebraic object called “rating map”. These are morphisms of commutative and idempotent monoids. We write such monoids $(R, +)$: the binary operation “+” is called *addition* and the neutral element is denoted by 0_R . Being idempotent means that $r + r = r$ for every $r \in R$. For every commutative and idempotent monoid $(R, +)$, one may define a canonical ordering \leq over R : for $r, s \in R$, we have $r \leq s$ when $r + s = s$. One may verify that \leq is a partial order which is compatible with addition.

► **Example 15.** For every set E , $(2^E, \cup)$ is an idempotent and commutative monoid. The neutral element is \emptyset and the canonical ordering is inclusion.

A rating map is a morphism $\rho : (2^{A^*}, \cup) \rightarrow (R, +)$ where $(R, +)$ is a *finite* idempotent and commutative monoid, called the *rating set of ρ* . That is, ρ is a map from 2^{A^*} to R such that $\rho(\emptyset) = 0_R$ and $\rho(K_1 \cup K_2) = \rho(K_1) + \rho(K_2)$ for every $K_1, K_2 \subseteq A^*$.

For the sake of improved readability, when applying a rating map ρ to a singleton set $\{w\}$, we write $\rho(w)$ for $\rho(\{w\})$. Moreover, we write $\rho_* : A^* \rightarrow R$ for the restriction of ρ to A^* : for every $w \in A^*$, we have $\rho_*(w) = \rho(w)$ (this notation is useful when referring to the language $\rho_*^{-1}(r) \subseteq A^*$, which consists of all words $w \in A^*$ such that $\rho(w) = r$).

Most of the theory makes sense for arbitrary rating maps. However, we shall often have to work with special rating maps satisfying additional properties. We define two kinds.

Nice rating maps. A rating map $\rho : 2^{A^*} \rightarrow R$ is *nice* when, for every nonempty language $K \subseteq A^*$, there exist finitely many words $w_1, \dots, w_n \in K$ such that $\rho(K) = \rho(w_1) + \dots + \rho(w_n)$.

When a rating map $\rho : 2^{A^*} \rightarrow R$ is nice, it is characterized by the canonical map $\rho_* : A^* \rightarrow R$. Indeed, for $K \subseteq A^*$, we may consider the sum of all elements $\rho(w)$ for $w \in K$: while it may be infinite, this sum boils down to a finite one since R is commutative and idempotent. The hypothesis that ρ is nice implies that $\rho(K)$ is equal to this sum.

Multiplicative rating maps. A rating map $\rho : 2^{A^*} \rightarrow R$ is multiplicative when its rating set R has more structure: it needs to be an *idempotent semiring*. A *semiring* is a tuple $(R, +, \cdot)$ where R is a set and “+” and “ \cdot ” are two binary operations called addition and multiplication. Moreover, $(R, +)$ is a commutative monoid, (R, \cdot) is a monoid (the neutral element is denoted by 1_R), the multiplication distributes over addition and the neutral element “ 0_R ” of $(R, +)$ is a zero for (R, \cdot) ($0_R \cdot r = r \cdot 0_R = 0_R$ for every $r \in R$). A semiring

R is *idempotent* when $r + r = r$ for every $r \in R$, *i.e.*, when the additive monoid $(R, +)$ is idempotent (there is no additional constraint on the multiplicative monoid (R, \cdot)).

► **Example 16.** A key example of an infinite idempotent semiring is the set 2^{A^*} . Union is the addition and language concatenation is the multiplication (with $\{\varepsilon\}$ as neutral element).

Let $\rho : 2^{A^*} \rightarrow R$ be a rating map: $(R, +)$ is an idempotent commutative monoid and ρ is a morphism from $(2^{A^*}, \cup)$ to $(R, +)$. We say that ρ is *multiplicative* when the rating set R is equipped with a multiplication “ \cdot ” such that $(R, +, \cdot)$ is an idempotent semiring and ρ is also a monoid morphism from $(2^{A^*}, \cdot)$ to (R, \cdot) . That is, the two following additional axioms have to be satisfied: $\rho(\varepsilon) = 1_R$ and $\rho(K_1 K_2) = \rho(K_1) \cdot \rho(K_2)$ for every $K_1, K_2 \subseteq A^*$.

► **Remark 17.** Rating maps which are both nice and multiplicative are finitely representable. As we explained, if $\rho : 2^{A^*} \rightarrow R$ is nice, it is characterized by the canonical map $\rho_* : A^* \rightarrow R$. When ρ is also multiplicative, ρ_* is finitely representable: it is a morphism into a finite monoid. Hence, we may speak of algorithms whose input is a nice multiplicative rating map.

Rating maps which are not nice and multiplicative cannot be finitely represented in general. Yet, they are crucial: while our main statements consider nice multiplicative rating maps, many proofs involve auxiliary rating maps which are neither nice nor multiplicative.

Optimal imprints. Now that we have rating maps, we turn to imprints. Consider a rating map $\rho : 2^{A^*} \rightarrow R$. Given any finite set of languages \mathbf{K} , we define the ρ -imprint of \mathbf{K} . Intuitively, when \mathbf{K} is a cover of some language L , this object measures the “quality” of \mathbf{K} . The ρ -imprint of \mathbf{K} is the following subset of R :

$$\mathcal{I}[\rho](\mathbf{K}) = \{r \mid r \leq \rho(K) \text{ for some } K \in \mathbf{K}\}.$$

We may now define optimality. Consider an arbitrary rating map $\rho : 2^{A^*} \rightarrow R$ and a Boolean algebra \mathcal{C} . Given a language L , an optimal \mathcal{C} -cover of L for ρ is a \mathcal{C} -cover \mathbf{K} of L which satisfies the following property:

$$\mathcal{I}[\rho](\mathbf{K}) \subseteq \mathcal{I}[\rho](\mathbf{K}') \quad \text{for every } \mathcal{C}\text{-cover } \mathbf{K}' \text{ of } L.$$

In general, there can be infinitely many optimal \mathcal{C} -covers for a given rating map ρ . It is shown in [22] that there always exists at least one (using closure under intersection for \mathcal{C}).

Clearly, for a Boolean algebra \mathcal{C} , a language L and a rating map ρ , all optimal \mathcal{C} -covers of L for ρ have the same ρ -imprint. Hence, this unique ρ -imprint is a *canonical* object for \mathcal{C} , L and ρ . We call it the *\mathcal{C} -optimal ρ -imprint on L* and we write it $\mathcal{I}_{\mathcal{C}}[L, \rho]$:

$$\mathcal{I}_{\mathcal{C}}[L, \rho] = \mathcal{I}[\rho](\mathbf{K}) \quad \text{for any optimal } \mathcal{C}\text{-cover } \mathbf{K} \text{ of } L \text{ for } \rho.$$

We complete the definition with a simple useful fact (a proof is available in [24]).

► **Fact 18.** Let \mathcal{C} be a Boolean algebra, $\rho : 2^{A^*} \rightarrow R$ a rating map and $L_1, L_2 \subseteq A^*$. Then, $\mathcal{I}_{\mathcal{C}}[L_1, \rho] \cup \mathcal{I}_{\mathcal{C}}[L_2, \rho] = \mathcal{I}_{\mathcal{C}}[L_1 \cup L_2, \rho]$.

Connection with covering. Consider the special case when the language L that needs to be covered is A^* . In that case, we write $\mathcal{I}_{\mathcal{C}}[\rho]$ for $\mathcal{I}_{\mathcal{C}}[A^*, \rho]$. It is shown in [22] that for every Boolean algebra \mathcal{C} , deciding \mathcal{C} -covering formally reduces to computing \mathcal{C} -optimal imprints from input nice multiplicative rating maps.

► **Proposition 19.** Let \mathcal{C} be a Boolean algebra. Assume that there exists an algorithm which computes $\mathcal{I}_{\mathcal{C}}[\rho]$ from an input nice multiplicative rating map ρ . Then, \mathcal{C} -covering is decidable.

4.2 Algorithm

We may now present our algorithm for $SF(\mathcal{C})$ -covering when \mathcal{C} is a finite quotient-closed Boolean algebra. We fix \mathcal{C} for the presentation. In view of Proposition 19, we need to prove that one may compute $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from an input nice multiplicative rating map ρ .

Our algorithm actually computes slightly more information. Since \mathcal{C} is a finite quotient-closed Boolean algebra, we may consider the equivalence $\sim_{\mathcal{C}}$ over A^* . In particular, the set $A^*/\sim_{\mathcal{C}}$ of $\sim_{\mathcal{C}}$ -classes is a finite monoid (we write “ \bullet ” for its multiplication) and the map $w \mapsto [w]_{\mathcal{C}}$ is a morphism. Given a rating map $\rho : 2^{A^*} \rightarrow R$ we define:

$$\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho] = \{(C, r) \in (A^*/\sim_{\mathcal{C}}) \times R \mid r \in \mathcal{I}_{SF(\mathcal{C})}[C, \rho]\}$$

Observe that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ captures more information than $\mathcal{I}_{SF(\mathcal{C})}[\rho]$. Indeed, it encodes all sets $\mathcal{I}_{SF(\mathcal{C})}[C, \rho]$ for $C \in A^*/\sim_{\mathcal{C}}$ and by Fact 18, $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ is the union of all these sets.

Our main result is a least fixpoint procedure for computing $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ from a nice multiplicative rating map ρ . It is based on a generic characterization theorem which we first present. Given an arbitrary nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$ and a set $S \subseteq (A^*/\sim_{\mathcal{C}}) \times R$, we say that S is $SF(\mathcal{C})$ -saturated for ρ when the following properties are satisfied:

1. **Trivial elements.** For every $w \in A^*$, we have $([w]_{\mathcal{C}}, \rho(w)) \in S$.
2. **Downset.** For every $(C, r) \in S$ and $q \in R$, if $q \leq r$, then $(C, q) \in S$.
3. **Multiplication.** For every $(C, q), (D, r) \in S$, we have $(C \bullet D, qr) \in S$.
4. **$SF(\mathcal{C})$ -closure.** For all $(E, r) \in S$, if $E \in A^*/\sim_{\mathcal{C}}$ is idempotent, then $(E, r^\omega + r^{\omega+1}) \in S$.

► **Theorem 20** (*$SF(\mathcal{C})$ -optimal imprints (\mathcal{C} finite)*). *Let $\rho : 2^{A^*} \rightarrow R$ be a nice multiplicative rating map. Then, $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ is the least $SF(\mathcal{C})$ -saturated subset of $(A^*/\sim_{\mathcal{C}}) \times R$ for ρ .*

Given a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$ as input, it is clear that one may compute the least $SF(\mathcal{C})$ -saturated subset of $(A^*/\sim_{\mathcal{C}}) \times R$ with a least fixpoint procedure. Hence, Theorem 20 provides an algorithm for computing $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$. As we explained above, we may then compute $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from this set. Together with Proposition 19, this yields Theorem 2 as a corollary: $SF(\mathcal{C})$ -covering is decidable when \mathcal{C} is a finite quotient-closed Boolean algebra. Theorem 20 is proved in the full version of this paper [25].

5 Covering when the input class is made of group languages

This section is devoted to Theorem 3. We show that when \mathcal{C} is a quotient-closed Boolean algebra of group languages with decidable separation, $SF(\mathcal{C})$ -covering is decidable.

As in Section 4, we rely on Proposition 19: we present an algorithm computing $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from an input nice multiplicative rating map ρ . We do not work with $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ itself but with another set carrying more information. Its definition requires introducing a few additional concepts. We first present them and then turn to the algorithm. For more details, see [26].

5.1 Preliminary definitions

Optimal ε -approximations. In this case, handling $SF(\mathcal{C})$ involves considering \mathcal{C} -optimal covers of $\{\varepsilon\}$. Since $\{\varepsilon\}$ is a singleton, there always exists such a cover consisting of a single language, which leads to the following definition.

Let \mathcal{C} be a Boolean algebra (we shall use the case when \mathcal{C} contains only group languages but this is not required for the definitions) and $\tau : 2^{A^*} \rightarrow Q$ a rating map. A \mathcal{C} -optimal ε -approximation for τ is a language $L \in \mathcal{C}$ such that $\varepsilon \in L$ and $\tau(L) \leq \tau(L')$ for every $L' \in \mathcal{C}$

satisfying $\varepsilon \in L'$. As expected, there always exists a \mathcal{C} -optimal ε -approximation for any rating map τ (see the full version of this paper [25] for a proof).

By definition, all \mathcal{C} -optimal ε -approximations for τ have the same image under τ . We write it $\mathfrak{i}_{\mathcal{C}}[\tau] \in Q$: $\mathfrak{i}_{\mathcal{C}}[\tau] = \tau(L)$ for every \mathcal{C} -optimal ε -approximation L for τ . It turns out that when τ is nice and multiplicative, computing $\mathfrak{i}_{\mathcal{C}}[\tau]$ from τ boils down to \mathcal{C} -separation. This is important: this is exactly how our algorithm for $SF(\mathcal{C})$ -covering depends on \mathcal{C} -separation.

► **Lemma 21.** *Let $\tau : 2^{A^*} \rightarrow Q$ be a nice rating map and \mathcal{C} a Boolean algebra. Then, $\mathfrak{i}_{\mathcal{C}}[\tau]$ is the sum of all $q \in Q$ such that $\{\varepsilon\}$ is not \mathcal{C} -separable from $\tau_*^{-1}(q)$.*

Nested rating maps. We want an algorithm which computes $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from an input nice multiplicative rating map ρ for a fixed quotient-closed Boolean algebra of group languages \mathcal{C} . Yet, we shall not use optimal ε -approximations with this input rating map ρ . Instead, we consider an auxiliary rating map built from ρ (the definition is taken from [24]).

Consider a Boolean algebra \mathcal{D} (we shall use the case $\mathcal{D} = SF(\mathcal{C})$) and a rating map $\rho : 2^{A^*} \rightarrow R$. We build a new map $\xi_{\mathcal{D}}[\rho] : 2^{A^*} \rightarrow 2^R$ whose rating set is $(2^R, \cup)$. For every $K \subseteq A^*$, we define $\xi_{\mathcal{D}}[\rho](K) = \mathcal{I}_{\mathcal{D}}[K, \rho] \in 2^R$. It follows from Fact 18 that this is indeed a rating map (on the other hand $\xi_{\mathcal{D}}[\rho]$ need not be nice nor multiplicative, see [24] for details).

We may now explain which set is computed by our algorithm instead of $\mathcal{I}_{SF(\mathcal{C})}[\rho]$. Consider a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. Since $\xi_{SF(\mathcal{C})}[\rho] : 2^{A^*} \rightarrow 2^R$ is a rating map, we may consider the element $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \in 2^R$. By definition, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] = \xi_{SF(\mathcal{C})}[\rho](L)$ where L is a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$. Therefore, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is a subset of $\xi_{SF(\mathcal{C})}[\rho](A^*) = \mathcal{I}_{SF(\mathcal{C})}[A^*, \rho] = \mathcal{I}_{SF(\mathcal{C})}[\rho]$. When \mathcal{C} is a quotient-closed Boolean algebra of group languages, one may compute the whole set $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from this subset.

► **Proposition 22.** *Let \mathcal{C} be a quotient-closed Boolean algebra of group languages and $\rho : 2^{A^*} \rightarrow R$ a nice multiplicative rating map. Then, $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ is the least subset of R containing $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ and satisfying the three following properties:*

- **Trivial elements.** *For every $w \in A$, $\rho(w) \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$.*
- **Downset.** *For every $r \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$ and $q \leq r$, we have $q \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$.*
- **Multiplication.** *For every $q, r \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$, we have $qr \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$.*

► **Remark 23.** Intuitively, we use $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ to “nest” two optimizations: one for \mathcal{C} and the other for $SF(\mathcal{C})$. Indeed, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] = \xi_{SF(\mathcal{C})}[\rho](L) = \mathcal{I}_{SF(\mathcal{C})}[L, \rho]$ where L is a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$. Hence, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is least set $\mathcal{I}[\rho](\mathbf{K}) \subseteq R$ (with respect to inclusion), over all $SF(\mathcal{C})$ -covers \mathbf{K} of some language $L \in \mathcal{C}$ containing ε .

5.2 Algorithm

We may now present our algorithm for computing $\mathcal{I}_{SF(\mathcal{C})}[\rho]$. We fix a quotient-closed Boolean algebra of group languages \mathcal{C} for the presentation. As expected, the main procedure computes $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ (see Proposition 22). In this case as well, this procedure is obtained from a characterization theorem.

Consider a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. We define the $SF(\mathcal{C})$ -complete subsets of R for ρ . The definition depends on auxiliary nice multiplicative rating maps. We first present them. Clearly, 2^R is an idempotent semiring (addition is union and the multiplication is lifted from the one of R). For every $S \subseteq R$, we use it as the rating set of a nice multiplicative rating map $\eta_{\rho, S} : 2^{A^*} \rightarrow 2^R$. Since we are defining a *nice* multiplicative rating map, it suffices to specify the evaluation of letters. For $a \in A$, we let $\eta_{\rho, S}(a) = S \cdot \{\rho(a)\} \cdot S \in 2^R$. Observe that by definition, we have $\mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}] \subseteq R$.

121:12 On all things star-free

We are ready to define the $SF(\mathcal{C})$ -complete subsets of R . Consider $S \subseteq R$. We say that S is $SF(\mathcal{C})$ -complete for ρ when the following conditions are satisfied:

1. **Downset.** For every $r \in S$ and $q \leq r$, we have $q \in S$.
2. **Multiplication.** For every $q, r \in S$, we have $qr \in S$.
3. **\mathcal{C} -operation.** We have $\mathfrak{i}_{\mathcal{C}}[\eta_{\rho,S}] \subseteq S$.
4. **$SF(\mathcal{C})$ -closure.** For every $r \in S$, we have $r^\omega + r^{\omega+1} \in S$.

► **Remark 24.** The definition of $SF(\mathcal{C})$ -complete subsets does not explicitly require that they contain some trivial elements. Yet, this is implied by \mathcal{C} -operation. Indeed, if $S \subseteq R$ is $SF(\mathcal{C})$ -complete, then $\eta_{\rho,S}(\varepsilon) = \{1_R\}$ (this is the multiplicative neutral element of 2^R). This implies that $1_R \in \mathfrak{i}_{\mathcal{C}}[\eta_{\rho,S}]$ and we obtain from \mathcal{C} -operation that $1_R \in S$.

► **Theorem 25** (*$SF(\mathcal{C})$ -optimal imprints (\mathcal{C} made of group languages)*). *Let $\rho : 2^{A^*} \rightarrow R$ be a nice multiplicative rating map. Then, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is the least $SF(\mathcal{C})$ -complete subset of R .*

When \mathcal{C} -separation is decidable, Theorem 25 yields a least fixpoint procedure for computing $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ from a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. The computation starts from the empty set and saturates it with the four operations in the definition of $SF(\mathcal{C})$ -complete subsets. It is clear that we may implement downset, multiplication and $SF(\mathcal{C})$ -closure. Moreover, we may implement \mathcal{C} -operation as this boils down to \mathcal{C} -separation by Lemma 21. Eventually, the computation reaches a fixpoint and it is straightforward to verify that this set is the least $SF(\mathcal{C})$ -complete subset of R , *i.e.*, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ by Theorem 25.

By Proposition 22, we may compute $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. Altogether, this yields the decidability of $SF(\mathcal{C})$ -covering by Proposition 19. Hence, Theorem 3 is proved.

6 Conclusion

We proved that for any quotient-closed Boolean algebra \mathcal{C} , $SF(\mathcal{C})$ -covering is decidable whenever \mathcal{C} is either finite or made of group languages and with decidable separation. Moreover, we presented an algebraic characterization of $SF(\mathcal{C})$ which holds for every quotient-closed Boolean algebra \mathcal{C} , generalizing earlier results [29, 16]. A key proof ingredient is an alternative definition of star-free closure: the operation $\mathcal{C} \mapsto SD(\mathcal{C})$ which we prove to be equivalent. This correspondence generalizes the work of Schützenberger [28] who introduced a single class SD (*i.e.* $SD(\{\emptyset, A^*\})$) corresponding to the star-free languages (*i.e.* $SF(\{\emptyset, A^*\})$).

Our results can be instantiated for several input classes \mathcal{C} . Theorem 2 applies when \mathcal{C} is finite. In this case, the only prominent application is the class of star-free languages itself. It was already known that covering is decidable for this class [9, 20]. However, Theorem 2 is important for two reasons. First, its proof is actually simpler than the earlier ones specific to the star-free languages (this is achieved by relying on the operation $\mathcal{C} \mapsto SD(\mathcal{C})$). More importantly, Theorem 2 is used as a key ingredient for proving our second generic statement: Theorem 3, which applies to classes made of group languages with decidable separation. It is known that separation is decidable for the class GR of all group languages [2]. Hence, we obtain that $SF(\text{GR})$ -covering is decidable. Another application is the class MOD consisting of languages counting the length of words modulo some number (deciding MOD-separation is a simple exercise). We get the decidability of $SF(\text{MOD})$ -covering. This is important, as the languages in $SF(\text{MOD})$ are those definable in first-order logic with modular predicates ($\text{FO}(<, \text{MOD})$). A last example is given by the input class consisting of all languages counting the number of occurrences of letters modulo some number. These are exactly the languages recognized by finite commutative groups, for which separation is decidable [5].

References

- 1 Jorge Almeida. Some algorithmic problems for pseudovarieties. *Publicationes Mathematicae Debrecen*, 54:531–552, 1999.
- 2 Christopher J. Ash. Inevitable graphs: a proof of the type II conjecture and some related decision procedures. *IJAC*, 1(1):127–146, 1991.
- 3 David A. Mix Barrington, Kevin Compton, Howard Straubing, and Denis Thérien. Regular languages in NC1. *Journal of Computer and System Sciences*, 44(3):478 – 499, 1992.
- 4 Janusz A. Brzozowski and Rina S. Cohen. Dot-depth of star-free events. *J. Comp. Sys. Sci.*, 5(1):1–16, 1971.
- 5 Manuel Delgado. Abelian pointlikes of a monoid. *Semigroup Forum*, 56:339–361, 1998.
- 6 Volker Diekert and Tobias Walter. Characterizing classes of regular languages using prefix codes of bounded synchronization delay. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, ICALP'16*, pages 129:1–129:14, 2016.
- 7 Volker Diekert and Tobias Walter. Characterizing classes of regular languages using prefix codes of bounded synchronization delay. *IJAC*, 27(6):561–590, 2017.
- 8 Samuel Eilenberg. *Automata, languages, and machines*, volume B. Academic Press, 1976.
- 9 Karsten Henckell. Pointlike sets: the finest aperiodic cover of a finite semigroup. *Journal of Pure Applied Algebra*, 55(1-2):85–126, 1988.
- 10 Joel Karnofsky and John Rhodes. Decidability of complexity one-half for finite semigroups. *Semigroup Forum*, 24(1):55–66, 1982. doi:10.1007/BF02572755.
- 11 Stuart W. Margolis and Jean-Éric Pin. Products of group languages. In *FCT'85*. Springer, 1985.
- 12 Robert McNaughton and Seymour A. Papert. *Counter-Free Automata*. MIT Press, 1971.
- 13 Jean-Éric Pin. Bridges for concatenation hierarchies. In *ICALP'98*, pages 431–442. Springer, 1998.
- 14 Jean-Éric Pin. The dot-depth hierarchy, 45 years later. In *The Role of Theory in Computer Science. Essays Dedicated to Janusz Brzozowski*. World Scientific Publishing, 2017.
- 15 Jean-Éric Pin. Mathematical foundations of automata theory. Lecture notes, 2019. URL: <https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf>.
- 16 Jean-Éric Pin, Howard Straubing, and Denis Thérien. Locally trivial categories and unambiguous concatenation. *Journal of Pure and Applied Algebra*, 52(3):297 – 311, 1988.
- 17 Thomas Place. Separating regular languages with two quantifiers alternations. In *LICS'15*, pages 202–213. IEEE Computer Society, 2015.
- 18 Thomas Place. Separating regular languages with two quantifier alternations. *Logical Methods in Computer Science*, 14(4), 2018.
- 19 Thomas Place and Marc Zeitoun. Going higher in the first-order quantifier alternation hierarchy on words. In *ICALP'14*, pages 342–353, 2014.
- 20 Thomas Place and Marc Zeitoun. Separating regular languages with first-order logic. *Logical Methods in Computer Science*, 12(1), 2016.
- 21 Thomas Place and Marc Zeitoun. Separation for dot-depth two. In *32th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS'17*, 2017. <https://arxiv.org/abs/1901.03361>.
- 22 Thomas Place and Marc Zeitoun. The covering problem. *Logical Methods in Computer Science*, 14(3), 2018.
- 23 Thomas Place and Marc Zeitoun. Generic results for concatenation hierarchies. *Theory of Computing Systems (ToCS)*, 2018. Selected papers from CSR'17.
- 24 Thomas Place and Marc Zeitoun. Separation for dot-depth two. In preparation, preprint at <http://www.labri.fr/~zeitoun/research/pdf/boolpol-full.pdf>, 2018.
- 25 Thomas Place and Marc Zeitoun. On all things star-free. Full version of this paper, <http://arxiv.org/abs/1904.11863>, 2019.
- 26 Thomas Place and Marc Zeitoun. Separation and covering for group based concatenation hierarchies. In *LICS'19*, 2019.

121:14 On all things star-free

- 27 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965.
- 28 Marcel Paul Schützenberger. Sur certaines opérations de fermeture dans les langages rationnels. *Symposia Mathematica*, XV:245–253, 1975. Convegno di Informatica Teorica, INDAM, Roma, 1973.
- 29 Howard Straubing. Aperiodic homomorphisms and the concatenation product of recognizable sets. *Journal of Pure and Applied Algebra*, 15(3):319 – 327, 1979.
- 30 Wolfgang Thomas. Classifying regular events in symbolic logic. *J. Comp. Sys. Sci.*, 25(3):360–376, 1982.
- 31 Thomas Wilke. Classifying discrete temporal properties. In *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science*, STACS’99, pages 32–46, 1999.

A Appendix to Section 3

This appendix presents the missing proofs in Section 3. It is written as a self-contained, full version of Section 3 which can be read independently.

We handle $SF(\mathcal{C})$ -membership with a generic algebraic characterization of $SF(\mathcal{C})$ (effective under the hypotheses of Theorems 2 and 3), generalizing earlier work by Pin, Straubing and Thérien [29, 16]. We rely on an alternate definition of star-free closure involving a semantic restriction of the Kleene star, which we first present.

A.1 Bounded synchronization delay

We define a second operation on classes of languages $\mathcal{C} \mapsto SD(\mathcal{C})$. We shall later prove that it coincides with star-free closure. It is based on the work of Schützenberger [28] who defined a single class SD corresponding to the star-free languages (*i.e.*, $SF(\{\emptyset, A^*\})$). Here, we generalize it as an operation. The definition involves a semantic restriction of the Kleene star operation on languages: it may only be applied to “*prefix codes with bounded synchronization delay*”. Introducing this notion requires basic definitions from code theory that we first recall.

A language $K \subseteq A^*$ is a *prefix code* when $\varepsilon \notin K$ and $K \cap KA^+ = \emptyset$ (no word in K has a strict prefix in K).

Given $d \geq 1$, a prefix code $K \subseteq A^+$ has *synchronization delay* d if for every $u, v, w \in A^*$ such that $uvw \in K^+$ and $v \in K^d$, we have $uv \in K^+$. Finally, a prefix code $K \subseteq A^+$ has *bounded synchronization delay* when it has synchronization delay d for some $d \geq 1$.

► **Example 26.** Let $A = \{a, b\}$. Clearly, $\{ab\}$ is a prefix code with synchronization delay 1: if $uvw \in (ab)^+$ and $v = ab$, we have $uv \in (ab)^+$. Similarly, one may verify that $(aab)^*ab$ is a prefix code with synchronization delay 2 (but not 1). On the other hand, $\{aa\}$ does not have bounded synchronization delay. If $d \geq 1$, $a(aa)^d a \in (aa)^*$ but $a(aa)^d \notin (aa)^*$.

We complement the definition with a few standard properties of prefix codes with bounded synchronization delay. They will be useful when manipulating them later. First, we have the following fact about prefix codes.

► **Fact 27.** Let K be a prefix code. Consider $m, n \in \mathbb{N}$, $u_1, \dots, u_m \in K$ and $v_1 \dots, v_n \in K$. The two following properties hold:

- If $u_1 \dots u_m$ is a prefix of $v_1 \dots v_n$, then $m \leq n$ and $u_i = v_i$ for every $i \leq m$.
- If $u_1 \dots u_m = v_1 \dots v_n$, then $m = n$ and $u_i = v_i$ for every $i \leq m$.

Proof. The second property is an immediate corollary of the first one. Hence, it suffices to show that if $u_1 \dots u_m$ is a prefix of $v_1 \dots v_n$, then $m \leq n$ and $u_i = v_i$ for every $i \leq m$. We proceed by induction on $m \in \mathbb{N}$. If $m = 0$, then $u_1 \dots u_m = \varepsilon$ and the property is immediate. Otherwise, $m \geq 1$. Clearly, $u_1 \dots u_{m-1}$ is a prefix of $v_1 \dots v_n$. Hence, induction yields that $m-1 \leq n$ and $u_i = v_i$ for every $i \leq m-1$. Therefore, since $u_1 \dots u_m$ is a prefix of $v_1 \dots v_n$, it follows that u_m is a prefix of $v_m \dots v_n$. Since $u_m \in K$ which is a prefix code, we have $u_m \neq \varepsilon$ which implies that $v_m \dots v_n \neq \varepsilon$, *i.e.* $n \geq m$. Moreover, since K is a prefix code an $u_m, v_m \in K$ we know that u_m is **not** a strict prefix of v_m and v_m is **not** a strict prefix of u_m . Together with the hypothesis that u_m is a prefix of $v_m \dots v_n$, this yields $u_m = v_m$, concluding the proof. ◀

The second assertion in Fact 27 implies that when K is a prefix code, every word $w \in K^*$ admits a *unique* decomposition witnessing this membership. This property is exactly the definition of “codes”, which are more general than prefix codes.

121:16 On all things star-free

We have the two following simple facts that we shall use to build new prefix codes with bounded synchronization delay out of already existing ones.

► **Fact 28.** *Let $d \geq 1$ and $K \subseteq A^+$ a prefix code with synchronization delay d . Then, every $H \subseteq K$ is also a prefix code with synchronization delay d .*

Proof. It is immediate from the definitions that H remains a prefix code. Hence, we need to prove that H has synchronization delay d . Consider $u, v, w \in A^*$ such that $uvw \in H^+$ and $v \in H^d$. We show that $uv \in H^+$. Since $H \subseteq K$, we have $uvw \in K^+$ and $v \in K^d$. Hence, since K has synchronization delay d , we obtain that $uv \in K^+$. Moreover, since K is a prefix code and $uvw \in K^+$, uvw admits a unique decomposition into factors of K (this follows from the second property in Fact 27). Additionally, since $uvw \in H^+$ with $H \subseteq K$, all factors in this unique decomposition belong to H . Finally, since $uv \in K^+$, the first property in Fact 27 yields that uv is a concatenation of factors in this unique decomposition. Hence, we have $uv \in H^+$ which concludes the proof. ◀

Additionally, we have the following more involved construction.

► **Fact 29.** *Let $d \geq 1$ and $K \subseteq A^+$ a prefix code with synchronization delay d . Let $H \subseteq K$. Then, $(K \setminus H)^*H$ is a prefix code with synchronization delay $d + 1$.*

Proof. We first verify that $(K \setminus H)^*H$ is a prefix code. Clearly, $(K \setminus H)^*H \subseteq A^+$ since $H \subseteq K \subseteq A^+$. Hence, we have to show that $(K \setminus H)^*HA^+ \cap (K \setminus H)^*H = \emptyset$. If w belongs to this intersection, in particular, we have $w \in K^*$. Therefore, since K is a prefix code, w admits a *unique* decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in K$. Since $w \in (K \setminus H)^*H$, the factor w_n is the only one in H . However, since $w \in (K \setminus H)^*HA^+$, the first property in Fact 27 implies that one of the factors w_i for $i \leq n - 1$ must belong to H . This is a contradiction.

It remains to show that $(K \setminus H)^*H$ has synchronization delay $d + 1$. We let $u, v, w \in A^*$ such that $uvw \in ((K \setminus H)^*H)^+$ and $v \in ((K \setminus H)^*H)^{d+1}$. We prove that $uv \in ((K \setminus H)^*H)^+$. Clearly $v = xy$ with $x \in ((K \setminus H)^*H)^d$ and $y \in (K \setminus H)^*H$. Observe that $x \in K^n$ for some $n \geq d$. Hence, since $uxyw = uvw \in K^+$ and K has synchronization delay d , it follows that $ux \in K^+$. Consequently $uv = uxy \in K^+(K \setminus H)^*H$. This implies that $uv \in K^*H$ which is clearly a subset of $((K \setminus H)^*H)^+$. This concludes the proof. ◀

We may now present the operation $\mathcal{C} \mapsto SD(\mathcal{C})$. The definition involves *unambiguous concatenation*. Given $K, L \subseteq A^*$, their concatenation KL is *unambiguous* when every word $w \in KL$ admits a *unique* decomposition $w = uv$ with $u \in K$ and $v \in L$. Given a class \mathcal{C} , $SD(\mathcal{C})$ is the least class containing \emptyset and $\{a\}$ for every $a \in A$, and closed under the following properties:

- **Intersection with \mathcal{C} :** if $K \in SD(\mathcal{C})$ and $L \in \mathcal{C}$, then $K \cap L \in SD(\mathcal{C})$.
- **Disjoint union:** if $K, L \in SD(\mathcal{C})$ are disjoint, then $K \uplus L \in SD(\mathcal{C})$.
- **Unambiguous product:** if $K, L \in SD(\mathcal{C})$ and KL is unambiguous, then $KL \in SD(\mathcal{C})$.
- **Kleene star for prefix codes with bounded synchronization delay:** if $K \in SD(\mathcal{C})$ is a prefix code with bounded synchronization delay, then $K^* \in SD(\mathcal{C})$.

► **Remark 30.** Schützenberger proved in [28] that $SD(\{\emptyset, A^*\}) = SF(\{\emptyset, A^*\})$. His definition of $SD(\{\emptyset, A^*\})$ was slightly less restrictive than ours: it does not require that the unions are disjoint and the concatenations unambiguous. It will be immediate from the correspondence with star-free closure that the two definitions are equivalent.

► **Remark 31.** This closure operation is different from standard ones. Instead of requiring that $\mathcal{C} \subseteq SD(\mathcal{C})$, we impose a stronger requirement: intersection with languages in \mathcal{C} is allowed. If we only required that $\mathcal{C} \subseteq SD(\mathcal{C})$, we would get a weaker operation which does not correspond to star-free closure in general. For example, let $A = \{a, b\}$ and consider the class MOD of Example 1. Observe $(aa)^* \in SD(\text{MOD})$. Indeed, $\{a\} \in SD(\text{MOD})$ has bounded synchronization delay, $(AA)^* \in \text{MOD}$ and $(aa)^* = a^* \cap (AA)^*$. Yet, one may verify that $(aa)^*$ cannot be built from the languages of MOD with union, concatenation and Kleene star applied to prefix codes with bounded synchronization delay.

A.2 Algebraic characterization of star-free closure

We now reduce deciding membership for $SF(\mathcal{C})$ to computing \mathcal{C} -stutters. Let us first define this new notion. Let \mathcal{C} be a quotient-closed Boolean algebra and $\alpha : A^* \rightarrow M$ be a morphism. A \mathcal{C} -stutter for α is an element $s \in M$ such that for every \mathcal{C} -cover \mathbf{K} of $\alpha^{-1}(s)$, there exists $K \in \mathbf{K}$ satisfying $K \cap KK \neq \emptyset$. When α is understood, we simply speak of \mathcal{C} -stutter. Finally, we say that α is \mathcal{C} -aperiodic when for every \mathcal{C} -stutter $s \in M$, we have $s^\omega = s^{\omega+1}$. The reduction is stated in the following theorem.

► **Theorem 7.** *Let \mathcal{C} be a quotient-closed Boolean algebra and consider a regular language $L \subseteq A^*$. The following properties are equivalent:*

1. $L \in SF(\mathcal{C})$.
2. $L \in SD(\mathcal{C})$.
3. *The syntactic morphism of L is \mathcal{C} -aperiodic.*

Naturally, the characterization need not be effective: this depends on \mathcal{C} . Deciding whether a morphism is \mathcal{C} -aperiodic boils down to computing \mathcal{C} -stutters. Yet, this is possible under the hypotheses of Theorems 2 and 3. First, if \mathcal{C} is a finite quotient-closed Boolean algebra, deciding whether an element is a \mathcal{C} -stutter is simple: there are finitely many \mathcal{C} -covers and we may check them all. Moreover, if \mathcal{C} is a quotient-closed Boolean algebra of group languages, the question boils down to \mathcal{C} -separation as stated in the next lemma.

► **Lemma 8.** *Let \mathcal{C} be a quotient-closed Boolean algebra of group languages and $\alpha : A^* \rightarrow M$ be a morphism. For all $s \in M$, s is a \mathcal{C} -stutter if and only if $\{\varepsilon\}$ is **not** \mathcal{C} -separable from $\alpha^{-1}(s)$.*

Proof. Assume first that s is a \mathcal{C} -stutter. We prove that $\{\varepsilon\}$ is not \mathcal{C} -separable from $\alpha^{-1}(s)$: given $L \in \mathcal{C}$ such that $\varepsilon \in L$, we show that $L \cap \alpha^{-1}(s) \neq \emptyset$. Let $\beta : A^* \rightarrow G$ be the syntactic morphism of L . Since \mathcal{C} is a quotient-closed Boolean algebra of group languages, it is standard and simple to verify that G is a group and that every language recognized by β belongs to \mathcal{C} (see [15] for example). Let \mathbf{K} be the set of all languages $\beta^{-1}(g)$ for $g \in G$ which intersect $\alpha^{-1}(s)$. By definition, \mathbf{K} is a \mathcal{C} -cover of $\alpha^{-1}(s)$. Since s is a \mathcal{C} -stutter, this yields $g \in G$ such that $\beta^{-1}(g)\beta^{-1}(g) \cap \beta^{-1}(g) \neq \emptyset$ and $\alpha^{-1}(s) \cap \beta^{-1}(g) \neq \emptyset$. Clearly, $\beta^{-1}(g)\beta^{-1}(g) \cap \beta^{-1}(g) \neq \emptyset$ implies that $gg = g$ which means that $g = 1_G$ since G is a group. Therefore, $\beta^{-1}(g) \subseteq L$ since L is recognized by its syntactic morphism β and $\varepsilon \in L$. Together with the hypothesis that $\alpha^{-1}(s) \cap \beta^{-1}(g) \neq \emptyset$, this implies $L \cap \alpha^{-1}(s) \neq \emptyset$ which concludes the proof for the first direction.

Conversely, we assume that $\{\varepsilon\}$ is not \mathcal{C} -separable from $\alpha^{-1}(s)$ and show that s is a \mathcal{C} -stutter. Let \mathbf{K} be a \mathcal{C} -cover of $\alpha^{-1}(s)$. We need to exhibit $K \in \mathbf{K}$ such that $K \cap KK \neq \emptyset$. Let H be the union of all languages in \mathbf{K} . Clearly, $H \in \mathcal{C}$ and $\alpha^{-1}(s) \subseteq H$. Therefore, since $\{\varepsilon\}$ is not \mathcal{C} -separable from $\alpha^{-1}(s)$, we have $\varepsilon \in H$. (otherwise $A^* \setminus H \in \mathcal{C}$ would be a separator). By definition of H , it follows that \mathbf{K} contains a language K such that $\varepsilon \in K$. It follows that $\varepsilon \in KK$ and we get $K \cap KK \neq \emptyset$ which concludes the proof. ◀

121:18 On all things star-free

Altogether, we obtain the membership part in Theorems 2 and 3: given a quotient-closed Boolean algebra which is either finite or made of group languages and with decidable separation, the $SF(\mathcal{C})$ -membership problem is decidable. It remains to prove Theorem 7.

A.3 Proof of Theorem 7

We prove Theorem 7. We fix an arbitrary quotient-closed Boolean algebra \mathcal{C} for the proof and show the implications $2 \Rightarrow 1) \Rightarrow 3) \Rightarrow 2)$.

Direction $2) \Rightarrow 1)$

We show that $SD(\mathcal{C}) \subseteq SF(\mathcal{C})$. This amounts to proving that $SF(\mathcal{C})$ satisfies all properties involved in the definition of $SD(\mathcal{C})$. In all cases but one, this is immediate. By definition of $SF(\mathcal{C})$, we have $\{a\} \in SF(\mathcal{C})$ for every $a \in A$. Moreover, $SF(\mathcal{C})$ is closed under union, intersection and concatenation by definition (this includes intersection with languages of \mathcal{C} since $\mathcal{C} \subseteq SF(\mathcal{C})$). Therefore, we concentrate on proving that $SF(\mathcal{C})$ is closed under Kleene star when it is applied to a prefix code with bounded synchronization delay. We fix $K \in SF(\mathcal{C})$ which is a prefix code with bounded synchronization delay and prove that $K^* \in SF(\mathcal{C})$. We let d be the delay and consider the following language H :

$$H = \left(A^* K^d \cap \left(A^* \setminus \left(A^* K^{d+1} \cup \bigcup_{0 \leq h \leq d} K^h \right) \right) \right) A^*$$

Since K has synchronization delay d , we have the following fact (note that this is the only part of the proof where we use the hypothesis that K has synchronization delay d).

► **Fact 32.** *We have $K^* \subseteq A^* \setminus H$.*

Proof. Clearly, we have, $K^* \subseteq A^* K^{d+1} \cup \bigcup_{0 \leq h \leq d} K^h$. Therefore, it follows that,

$$A^* \setminus \left(A^* K^{d+1} \cup \bigcup_{0 \leq h \leq d} K^h \right) \subseteq A^* \setminus K^*$$

By definition of H , this yields $H \subseteq (A^* K^d \cap (A^* \setminus K^*)) A^*$ and we obtain,

$$A^* \setminus ((A^* K^d \cap (A^* \setminus K^*)) A^*) \subseteq A^* \setminus H$$

Moreover, since K has synchronization delay d , one may verify that,

$$K^* \subseteq A^* \setminus ((A^* K^d \cap (A^* \setminus K^*)) A^*)$$

Altogether, this yields as desired that $K^* \subseteq A^* \setminus H$. ◀

Clearly, we have $H \in SF(\mathcal{C})$ by definition of $SF(\mathcal{C})$ since $K, A^* \in SF(\mathcal{C})$. Finally, we define,

$$G = \left(\bigcup_{0 \leq h \leq d-1} K^h \right) \cup (A^* K^d \cap (A^* \setminus H))$$

Again, it is immediate that $G \in SF(\mathcal{C})$. We show that $K^* = G$ which concludes the proof. Let us start with the left to right inclusion. Consider $x \in K^*$. If $x \in K^h$ for $h \leq d-1$, it is

immediate that $w \in G$ by definition. Otherwise, we have $x \in A^*K^d$ and since $x \in K^*$, we know that $x \in A^* \setminus H$ by Fact 32 (note that this fact relies on the hypothesis that K has bounded synchronization delay). This implies that $x \in A^*K^d \cap (A^* \setminus H)$ which yields $x \in G$, finishing the proof for this inclusion.

We finish with the right to left inclusion (which is independent from the hypothesis that K has bounded synchronization delay). Consider $x \in G$, we show that $x \in K^*$. If $x \in \bigcup_{0 \leq h \leq d-1} K^h$, this is immediate. Otherwise, $x \in A^*K^d \cap (A^* \setminus H)$ and we proceed by induction on the length of x . By hypothesis, $x \in A^*K^d$ and $x \notin H$. By definition of H , this implies that,

$$x \in A^*K^{d+1} \cup \bigcup_{0 \leq h \leq d} K^h$$

If $x \in \bigcup_{0 \leq h \leq d} K^h$, it is immediate that $x \in K^*$ and we are finished. Otherwise, $x \in A^*K^{d+1}$ which means that $x = x'y$ with $x' \in A^*K^d$ and $y \in K$. Since $\varepsilon \notin K$ (K is a prefix code), we have $y \neq \varepsilon$ which implies that $|x'| < |x|$. Moreover, since $x \in A^* \setminus H$ and x' is a prefix of x , one may verify from the definition of H that $x' \in A^* \setminus H$ as well. Altogether, we have $x' \in A^*K^d \cap (A^* \setminus H)$ which satisfies $|x'| < |x|$. Therefore, induction yields that $x' \in K^*$. Finally, since $y \in K$, we get $x = x'y \in K^*K \subseteq K^*$ which concludes the proof.

Direction 1) \Rightarrow 3)

Consider a regular language $L \in SF(\mathcal{C})$ and let $\alpha : A^* \rightarrow M$ be its syntactic morphism. We prove that α is \mathcal{C} -aperiodic. First, we show that one may assume without loss of generality that \mathcal{C} is *finite*.

► **Fact 33.** *There exists a finite quotient-closed Boolean algebra \mathcal{D} such that $\mathcal{D} \subseteq \mathcal{C}$ and $L \in SF(\mathcal{D})$.*

Proof. Since $L \in SF(\mathcal{C})$, it is built from finitely many languages in \mathcal{C} using the operations available in star-free closure. We let \mathcal{D} be the least quotient-closed Boolean algebra containing these languages. One may verify that \mathcal{D} is finite (this is because a regular language has finitely many quotients by the Myhill-Nerode theorem, see Lemma 17 in [23] for a proof). Moreover, it is immediate that $\mathcal{D} \subseteq \mathcal{C}$ and $L \in SF(\mathcal{D})$ by definition. ◀

We fix the finite quotient-closed Boolean algebra \mathcal{D} described in Fact 33. Moreover, we consider the associated canonical equivalence $\sim_{\mathcal{D}}$ over A^* . We have the following lemma.

► **Lemma 34.** *Consider $K \in SF(\mathcal{D})$. There exists $k \in \mathbb{N}$ such that for every $\ell \geq k$ and $u, x, y \in A^*$ satisfying $uu \sim_{\mathcal{D}} u$, we have $xu^\ell y \in K$ if and only if $xu^\ell uy \in K$.*

Proof. By definition, K is built from languages in \mathcal{D} and the singletons $\{a\}$ for $a \in A$ using union, complement and concatenation. We proceed by induction on this construction.

Assume first that $K \in \mathcal{D}$. In that case, the proposition holds for $k = 1$. Indeed, if $uu \sim_{\mathcal{C}} u$, then $xu^\ell uy \sim_{\mathcal{C}} xu^\ell y$ for every $\ell \geq 1$ and $x, y \in A^*$ since $\sim_{\mathcal{C}}$ is a congruence. Therefore, since $K \in \mathcal{D}$, it follows that $xu^\ell y \in K$ if and only if $xu^\ell uy \in K$ by definition of $\sim_{\mathcal{C}}$. If $K = \{a\}$, one may verify that the property holds for $k = 2$.

We turn to the inductive cases. Assume first that the last operation used to build K is union. We have $K = K_1 \cup K_2$ where $K_1, K_2 \in SF(\mathcal{D})$ are simpler languages. Induction yields $k_1, k_2 \in \mathbb{N}$ such that for $i = 1, 2$, if $\ell \geq k_i$ and $u, x, y \in A^*$ satisfy $uu \sim_{\mathcal{C}} u$, we have $xu^\ell y \in K_i$ if and only if $xu^\ell uy \in K_i$. It is immediate that in this case, the proposition holds

for k as the maximum between k_1 and k_2 . We turn to complement. Assume that $K = A^* \setminus H$ where $H \in SF(\mathcal{D})$ is a simpler language. Induction yields $h \in \mathbb{N}$ such that, if $\ell \geq h$ and $u, x, y \in A^*$ satisfy $uu \sim_{\mathcal{C}} u$, we have $xu^\ell y \in H$ if and only if $xu^\ell uy \in H$. It is immediate that in this case, the proposition holds for $k = h$.

Finally, we assume that the last operation used to construct K is concatenation. We have $K = K_1K_2$ with $K_1, K_2 \in SF(\mathcal{D})$ are simpler languages. Induction yields $k_1, k_2 \in \mathbb{N}$ such that for $i = 1, 2$, if $\ell \geq k_i$ and $u, x, y \in A^*$ satisfy $uu \sim_{\mathcal{C}} u$, we have $xu^\ell y \in K_i$ if and only if $xu^\ell uy \in K_i$. Let m be the maximum between k_1 and k_2 . We prove that the proposition holds for $k = 2m + 1$. Let $\ell \geq k$ and $u, x, y \in A^*$ such that $uu \sim_{\mathcal{C}} u$. We need to show that $xu^\ell y \in K$ if and only if $xu^\ell uy \in K$. We concentrate on the right to left direction (the converse one is symmetrical). Assuming that $xu^\ell uy \in K$, we show that $xu^\ell y \in K$. Since $K = K_1K_2$, we get $w_1 \in K_1$ and $w_2 \in K_2$ such that $xu^\ell uy = w_1w_2$. Since $\ell \geq 2m + 1$, it follows that either $xu^m u$ is a prefix of w_1 or $u^m uy$ is a suffix of w_2 . By symmetry we assume that the former property holds: we have $w_1 = xu^m uz$ for some $z \in A^*$. Observe that since $xu^\ell uy = w_1w_2$, it follows that $zw_2 = u^{\ell-m}y$. Moreover, we have $m \geq k_1$ by definition of m . Therefore, since $xu^m uz = w_1 \in K_1$, we know that $xu^m z \in K_1$ by definition of k_1 . Thus, $xu^m zw_2 \in K_1K_2 = K$. Since $zw_2 = u^{\ell-m}y$, this yields $xu^\ell y \in K$, concluding the proof. \blacktriangleleft

We are ready to prove that the syntactic morphism $\alpha : A^* \rightarrow M$ of L is \mathcal{D} -aperiodic. We fix a \mathcal{C} -stutter $s \in M$ for the proof and show that $s^\omega = s^{\omega+1}$. Let \mathbf{K} be the set containing all $\sim_{\mathcal{D}}$ -classes intersecting $\alpha^{-1}(s)$. Clearly, \mathbf{K} is a \mathcal{D} -cover of $\alpha^{-1}(s)$ and therefore a \mathcal{C} -cover as well since $\mathcal{D} \subseteq \mathcal{C}$. Hence, since s is a \mathcal{C} -stutter, there exists $K \in \mathbf{K}$ such that $K \cap KK \neq \emptyset$.

By definition K is a $\sim_{\mathcal{D}}$ -class. Moreover, $\sim_{\mathcal{D}}$ is a congruence for concatenation. Therefore, $K \cap KK \neq \emptyset$ implies that $KK \subseteq K$. Moreover, since $\alpha^{-1}(s) \cap K \neq \emptyset$ by definition of \mathbf{K} , there exists some word $u \in \alpha^{-1}(s) \cap K$. We have $\alpha(u) = s$. Since K is a $\sim_{\mathcal{D}}$ -class such that $KK \subseteq K$, we have $uu \sim_{\mathcal{D}} u$. Therefore, since $K \in SF(\mathcal{D})$, Lemma 34 yields a number $k \in \mathbb{N}$ such that for every $x, y \in A^*$ and every $\ell \geq k$, we have $xu^\ell y \in L$ if and only if $xu^\ell uy \in L$. In particular, this holds for $\ell = k\omega$ (where ω is the idempotent power of M). Hence, we have $xu^{k\omega} y \in L$ if and only if $xu^{k\omega} uy \in L$. This exactly says that $u^{k\omega}$ and $u^{k\omega}u$ are equivalent for the syntactic congruence of L . Consequently, since α is the syntactic morphism of L , we have $\alpha(u^{k\omega}) = \alpha(u^{k\omega}u)$. Since $\alpha(u) = s$, this exactly says that $s^\omega = s^{\omega+1}$, concluding the proof.

Direction 3) \Rightarrow 2)

Consider a \mathcal{C} -aperiodic morphism $\alpha : A^* \rightarrow M$. We show that every language recognized by α belongs to $SD(\mathcal{C})$. We start with two definitions.

Given $K \subseteq A^*$ and $s \in M$, we say that K is s -safe when $s\alpha(u) = s\alpha(v)$ for every $u, v \in K$. We extend this notion to sets of languages: such a set \mathbf{K} is s -safe when every $K \in \mathbf{K}$ is s -safe. Finally, given a language $P \subseteq A^*$, an $SD(\mathcal{C})$ -partition of P is a finite partition of P into languages of $SD(\mathcal{C})$. The argument is based on the following proposition.

► Proposition 9. *Let $P \subseteq A^+$ be a prefix code with bounded synchronization delay. Assume that there exists a 1_M -safe $SD(\mathcal{C})$ -partition of P . Then, for every $s \in M$, there exists an s -safe $SD(\mathcal{C})$ -partition of P^* .*

We first apply Proposition 9 to conclude the main argument. We show that every language recognized by α belongs to $SD(\mathcal{C})$. By definition, $SD(\mathcal{C})$ is closed under disjoint union. Hence, it suffices to show that $\alpha^{-1}(t) \in SD(\mathcal{C})$ for every $t \in M$. We fix $t \in M$ for the proof.

Clearly, $A \subseteq A^+$ is a prefix code with bounded synchronization delay and $\{\{a\} \mid a \in A\}$ is a 1_M -safe $SD(\mathcal{C})$ -partition of A . Hence, Proposition 9 (applied in the case $s = 1_M$) yields a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{K} of A^* . One may verify that $\alpha^{-1}(t)$ is the disjoint union of all $K \in \mathbf{K}$ intersecting $\alpha^{-1}(t)$. Hence, $\alpha^{-1}(t) \in SD(\mathcal{C})$ which concludes the main argument.

It remains to prove Proposition 9. We let $P \subseteq A^*$ be a prefix code with bounded synchronization delay and \mathbf{H} be a 1_M -safe $SD(\mathcal{C})$ -partition of P . Moreover, we fix $s \in M$. We need to build an $SD(\mathcal{C})$ -partition \mathbf{K} of P^* such that every $K \in \mathbf{K}$ is s -safe. We proceed by induction on the three following parameters listed by order of importance: (1) the size of $\alpha(P^+) \subseteq M$, (2) the size of \mathbf{H} and (3) the size of $s \cdot \alpha(P^*) \subseteq M$.

We distinguish two cases depending on the following property of s and \mathbf{H} . We say that s is \mathbf{H} -stable when the following holds:

$$\text{for every } H \in \mathbf{H}, \quad s \cdot \alpha(P^*) = s \cdot \alpha(P^*H). \quad (3)$$

The base case happens when s is \mathbf{H} -stable. Otherwise, we use induction on our parameters.

Base case: s is \mathbf{H} -stable. Since α is \mathcal{C} -aperiodic, we have the following simple fact.

► **Fact 10.** *There is a finite quotient-closed Boolean algebra $\mathcal{D} \subseteq \mathcal{C}$ such that α is \mathcal{D} -aperiodic.*

Proof. By definition, for every $s \in M$ which is **not** a \mathcal{C} -stutter, there exists a \mathcal{C} -cover \mathbf{K}_s of $\alpha^{-1}(s)$ such that $K \cap KK = \emptyset$ for every $K \in \mathbf{K}_s$. Let \mathbf{H} be the union of all sets \mathbf{K}_s . Since \mathbf{H} is a finite set of languages in \mathcal{C} , one may verify that there exists a finite quotient-closed Boolean algebra $\mathcal{D} \subseteq \mathcal{C}$ containing every $H \in \mathbf{H}$ (see Lemma 17 in [23] for a proof). It is now immediate that α is \mathcal{D} -aperiodic. ◀

Since \mathcal{D} is finite, we may consider the associated canonical equivalence $\sim_{\mathcal{D}}$ over A^* . We let $\mathbf{K} = \{P^* \cap D \mid D \in A^*/\sim_{\mathcal{D}}\}$. Clearly, \mathbf{K} is a partition of P^* . Let us verify that it only contains languages in $SD(\mathcal{C})$. We have $P \in SD(\mathcal{C})$: it is the disjoint union of all languages in the $SD(\mathcal{C})$ -partition \mathbf{H} of P . Moreover, $P^* \in SD(\mathcal{C})$ since P is a prefix code with bounded synchronization delay. Hence, $P^* \cap D \in SD(\mathcal{C})$ for every $D \in A^*/\sim_{\mathcal{D}}$ since $D \in \mathcal{D} \subseteq \mathcal{C}$. Therefore, it remains to show that every language $K \in \mathbf{K}$ is s -safe. This is a consequence of the following lemma which is proved using the hypothesis (3) that s is \mathbf{H} -stable.

► **Lemma 11.** *For every $u, v \in P^*$ such that $u \sim_{\mathcal{D}} v$, we have $s\alpha(u) = s\alpha(v)$.*

Proof. We first use the hypothesis that s is \mathbf{H} -stable to prove the following fact.

► **Fact 35.** *Let $q, e \in \alpha(P^*)$ such that e is idempotent. Then, we have $sqe = sq$.*

Proof. The proof is based on the following preliminary result. For every $x, y \in P^*$, we show that there exists $r \in \alpha(P^*)$ such that $sr\alpha(x) = s\alpha(y)$. Since $x \in P^*$, there exists a decomposition $x = x_1 \cdots x_n$ with $x_1, \dots, x_n \in P$. We proceed by induction on the length n of this decomposition. If $n = 0$, then $x = \varepsilon$ and it suffices to choose $r = \alpha(y) \in \alpha(P^*)$. Otherwise, $x = wx'$ with $w \in P$, $x' \in P^*$ admitting a decomposition of length $n - 1$. Induction yields $r' \in \alpha(P^*)$ such that $sr'\alpha(x') = s\alpha(y)$. Moreover, since $w \in P$ and \mathbf{H} is a partition of P , there exists some $H \in \mathbf{H}$ such that $w \in H$. Since $r' \in \alpha(P^*)$, we may then use the hypothesis that s is \mathbf{H} -stable to obtain $r \in \alpha(P^*)$ and $t \in \alpha(H)$ such that $sr' = srt$. Finally, we know that \mathbf{H} is 1_M -safe by hypothesis. Hence, since $t, \alpha(w) \in \alpha(H)$, we have $t = \alpha(w)$ and $sr' = sr\alpha(w)$. Altogether, this yields $sr\alpha(x) = sr\alpha(w)\alpha(x') = sr'\alpha(x') = s\alpha(y)$ which concludes the proof of our preliminary result.

It remains to prove the fact. Consider $q, e \in \alpha(P^*)$ such that e is idempotent. By definition, we have $x, y \in P^*$ such that $q = \alpha(y)$ and $e = \alpha(x)$. Hence, we have $r \in \alpha(P^*)$

121:22 On all things star-free

such that $sre = sq$. Finally, since e is idempotent, we obtain that $sqe = sree = sre = sq$ which completes the proof. \blacktriangleleft

We may now prove the lemma. Consider $u, v \in P^*$ such that $u \sim_{\mathcal{D}} v$. We show that $s\alpha(u) = s\alpha(v)$. We first consider the special case when $u \sim_{\mathcal{D}} uu$.

Assume that $u \sim_{\mathcal{D}} uu$. We show that $s\alpha(u)$ and $s\alpha(v)$ are both equal to s . Since $\sim_{\mathcal{D}}$ is a congruence and $u \sim_{\mathcal{D}} v$, we also have $v \sim_{\mathcal{D}} vv$. Hence, it suffices to use the hypothesis that $u \sim_{\mathcal{D}} uu$ to show $s\alpha(u) = s$ (the same result is obtained for v by symmetry). Observe that $\alpha(u)$ must be a \mathcal{D} -stutter. Indeed, if \mathbf{K} is a \mathcal{D} -cover of $\alpha^{-1}(\alpha(u))$, there exists $K \in \mathbf{K}$ such that $u \in K$. Hence, $uu \in K$ since $K \in \mathcal{D}$ and $u \sim_{\mathcal{D}} uu$. Thus, $K \in \mathbf{K}$ satisfies $K \cap KK \neq \emptyset$ (uu is in the intersection). By Fact 10, α is \mathcal{D} -aperiodic. Therefore, $\alpha(u)$ being a \mathcal{D} -stutter implies that $(\alpha(u))^\omega = (\alpha(u))^\omega \alpha(u)$. We may multiply by s on the left to get $s(\alpha(u))^\omega = s(\alpha(u))^\omega \alpha(u)$. Moreover, since $(\alpha(u))^\omega$ is an idempotent of $\alpha(P^*)$, it follows from Fact 35 that $s = s(\alpha(u))^\omega$. Altogether, this yields $s\alpha(u) = s$, concluding this case.

It remains to handle the case when u is not necessarily equivalent to uu for $\sim_{\mathcal{D}}$. Since $\sim_{\mathcal{D}}$ is a congruence, the quotient set $A^*/\sim_{\mathcal{D}}$ is a finite monoid and it is standard that there exists a number $p \geq 1$ such that $u^p \sim_{\mathcal{D}} u^p u^p$ (i.e. the $\sim_{\mathcal{D}}$ -class of u^p is an idempotent) and $\alpha(u^p) \in \alpha(P^*)$ is idempotent. Moreover, since $u \sim_{\mathcal{D}} v$, we have $u^p \sim_{\mathcal{D}} v u^{p-1}$. Hence, since $u^p \sim_{\mathcal{D}} u^p u^p$, it follows from the special case treated above that $s\alpha(u^p) = s\alpha(v u^{p-1})$. Moreover, we may multiply by $\alpha(u)$ on the right side which yields, $s\alpha(u)\alpha(u^p) = s\alpha(v)\alpha(u^p)$. Finally, since $\alpha(u^p)$ is idempotent, it follows from Fact 35 that $s\alpha(u)\alpha(u^p) = s\alpha(u)$ and $s\alpha(v)\alpha(u^p) = s\alpha(v)$. Altogether, we obtain that $s\alpha(u) = s\alpha(v)$, concluding the proof. \blacktriangleleft

Inductive step: s is not \mathbf{H} -stable. By hypothesis, we know that (3) does not hold. Therefore, we get some $H \in \mathbf{H}$ such that the following **strict** inclusion holds,

$$s \cdot \alpha(P^*H) \subsetneq s \cdot \alpha(P^*). \quad (4)$$

We fix this language $H \in \mathbf{H}$ for the remainder of the proof. The following lemma is proved by induction on our second parameter (the size of \mathbf{H}).

► **Lemma 12.** *There exists a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{U} of $(P \setminus H)^*$.*

Proof. By Fact 28, $P \setminus H$ remains a prefix code with bounded synchronization delay since it is included in P . Moreover, it is immediate that $\mathbf{G} = \mathbf{H} \setminus \{H\}$ is a $SD(\mathcal{C})$ -partition of $P \setminus H$ such that every $G \in \mathbf{G}$ is 1_M -safe. Additionally, it is clear that $\alpha((P \setminus H)^+) \subseteq \alpha(P^+)$ (our first induction parameter has not increased) and $\mathbf{G} \subsetneq \mathbf{H}$ (our second parameter has decreased). Hence, we may apply induction in Proposition 9 for the case when P, \mathbf{H} and s have been replaced by $P \setminus H, \mathbf{G}$ and 1_M . This yields a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{U} of $(P \setminus H)^*$. \blacktriangleleft

We fix the partition \mathbf{U} of $(P \setminus H)^*$ given by Lemma 12 and distinguish two independent sub-cases. Since $H \subseteq P$ (as H is an element of the partition \mathbf{H} of P), we have $\alpha(P^*H) \subseteq \alpha(P^+)$. We use a different argument depending on whether this inclusion is strict or not.

Sub-case 1: $\alpha(P^*H) = \alpha(P^+)$. Since H is 1_M -safe by hypothesis, there exists $t \in M$ such that $\alpha(H) = \{t\}$. Similarly, since every $U \in \mathbf{U}$ is 1_M -safe, there exists $r_U \in M$ such that $\alpha(U) = \{r_U\}$. The construction of \mathbf{K} is based on the next lemma which is proved using (2), the hypothesis of Sub-case 1 and induction on our third parameter (the size of $s \cdot \alpha(P^*) \subseteq M$).

► **Lemma 13.** *For every $U \in \mathbf{U}$, there exists an $sr_U t$ -safe $SD(\mathcal{C})$ -partition \mathbf{W}_U of P^* .*

Proof. We fix $U \in \mathbf{U}$ for the proof. Since \mathbf{U} is a partition of $(P \setminus H)^*$, we have $\alpha(U) \subseteq \alpha(P^*)$ which means that $r_U \in \alpha(P^*)$. Thus, we have $sr_U t \in s\alpha(P^*H)$. Therefore, $sr_U t\alpha(P^*) \subseteq s\alpha(P^*HP^*)$ and since $H \subseteq P$, we get $sr_U t\alpha(P^*) \subseteq s\alpha(P^+)$. Combined with our hypothesis in Sub-case 1 (i.e. $\alpha(P^*H) = \alpha(P^+)$), this yields $sr_U t\alpha(P^*) \subseteq s\alpha(P^*H)$. Finally, we obtain from (4) (i.e. $s\alpha(P^*H) \subsetneq s\alpha(P^*)$) that the **strict** inclusion $sr_U t\alpha(P^*) \subsetneq s\alpha(P^*)$ holds. Consequently, by induction on our third parameter (i.e. the size of $s\alpha(P^*)$) we may apply Proposition 9 in the case when $s \in M$ has been replaced by $sr_U t \in M$. Note that here, our first two parameters have not increased (they only depend on P and \mathbf{H} which remain unchanged). This yields the desired $SF(\mathcal{C})$ -partition \mathbf{W}_U of P^* . ◀

We are ready to define the partition \mathbf{K} of P^* . Using Lemma 13, we define,

$$\mathbf{K} = \mathbf{U} \cup \bigcup_{U \in \mathbf{U}} \{UHW \mid W \in \mathbf{W}_U\}$$

It remains to show that \mathbf{K} is an s -safe $SD(\mathcal{C})$ -partition of P^* . First, \mathbf{K} is a partition of P^* since P is a prefix code and $H \subseteq P$. Indeed, every word $w \in P^*$ admits a *unique* decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in P$. If no factor w_i belongs to H , then $w \in (P \setminus H)^*$ and w belongs to some unique $U \in \mathbf{U}$. Otherwise, let w_i be the leftmost factor such that $w_i \in H$. Thus, $w_1 \cdots w_{i-1} \in (P \setminus H)^*$, which also yields a unique $U \in \mathbf{U}$ such that $w_1 \cdots w_{i-1} \in U$ and $w_{i+1} \cdots w_n \in P^*$ which yields a unique $W \in \mathbf{W}_U$ such that $w_{i+1} \cdots w_n \in W$. It follows that $w \in UHW$ which is an element of \mathbf{K} (and the only one containing w).

Moreover, every $K \in \mathbf{K}$ belongs to $SD(\mathcal{C})$. If $K \in \mathbf{U}$, this is immediate by definition of \mathbf{U} in Lemma 12. Otherwise, $K = UHW$ with $U \in \mathbf{U}$ and $W \in \mathbf{W}_U$. We know that $U, H, W \in SD(\mathcal{C})$ by definition. Moreover, one may verify that the concatenation UHW is *unambiguous* since P is a prefix code, $U \subseteq (P \setminus H)^*$ and $W \subseteq H^*$. Hence, $K \in SD(\mathcal{C})$.

Finally, we verify that \mathbf{K} is s -safe. Consider $K \in \mathbf{K}$ and $w, w' \in K$, we show that $s\alpha(w) = s\alpha(w')$. If $K \in \mathbf{U}$, this is immediate: \mathbf{U} is 1_M -safe by definition. Otherwise, $K = UHW$ with $U \in \mathbf{U}$ and $W \in \mathbf{W}_U$. By definition, $\alpha(H) = \{t\}$ and $\alpha(U) = \{r_U\}$ which implies that $s\alpha(w) = str_U \alpha(x)$ and $s\alpha(w') = str_U \alpha(x')$ for $x, x' \in W$. Moreover, $W \in \mathbf{W}_U$ is $sr_U t$ -safe by definition. Hence, $s\alpha(w) = s\alpha(w')$, which concludes the proof of this sub-case.

Sub-case 2: we have the strict inclusion $\alpha(P^*H) \subsetneq \alpha(P^+)$. Consider $w \in P^*$. Since P is a prefix code, w admits a unique decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in P$. We may look at the rightmost factor $w_i \in H \subseteq P$ to uniquely decompose w in two parts (each of them possibly empty): the prefix $w_1 \cdots w_i \in ((P \setminus H)^*H)^*$ and the suffix in $w_{i+1} \cdots w_n \in (P \setminus H)^*$. Using induction, we construct $SD(\mathcal{C})$ -partitions of the possible languages of prefixes and suffixes. Then, we combine them to construct a partition of the whole set P^* . We already handled the suffixes: \mathbf{H} is an $SD(\mathcal{C})$ -partition of $(P \setminus H)^*$. The prefixes are handled with the next lemma, whose proof uses the hypothesis of Sub-case 2 and induction on our first parameter (the size of $\alpha(P^+)$).

► **Lemma 14.** *There exists a 1_M -safe $SD(\mathcal{C})$ -partition \mathbf{V} of $((P \setminus H)^*H)^*$.*

Proof. Let $Q = (P \setminus H)^*H$. By Fact 29, Q remains a prefix code with bounded synchronization delay. We apply induction in Proposition 9 for the case when P has been replaced by Q . Doing so requires building an appropriate $SF(\mathcal{C})$ -partition of Q and proving that one of our induction parameters has decreased.

Let $\mathbf{F} = \{UH \mid U \in \mathbf{U}\}$. Since \mathbf{U} is a partition of $(P \setminus H)^*$ and P is a prefix code, one may verify that \mathbf{F} is a partition of $Q = (P \setminus H)^*H$. Moreover, it only contains languages in

$SD(\mathcal{C})$. Indeed, if $U \in \mathbf{U}$, then the concatenation UH is *unambiguous* since $U \subseteq (P \setminus H)^*$ and P is a prefix code. Moreover, $U, H \in SD(\mathcal{C})$ by hypothesis. Finally, UH is 1_M -safe since this is the case for both U and H by definition. It remains to show that our induction parameters have decreased. Since $Q = (P \setminus H)^*H$, it is clear that $Q^+ \subseteq P^*H$. Hence, $\alpha(P^*H) \subsetneq \alpha(P^+)$ by hypothesis in Sub-case 2, we have $\alpha(Q^+) \subsetneq \alpha(P^+)$: our first induction parameter has decreased. Thus, we may apply Proposition 9 in the case when P, \mathbf{H} and s have been replaced by Q, \mathbf{F} and 1_M . This yields the desired $SD(\mathcal{C})$ -partition \mathbf{V} of $((P \setminus H)^*H)^*$. ◀

Using Lemma 14, we define $\mathbf{K} = \{VU \mid V \in \mathbf{V} \text{ and } U \in \mathbf{U}\}$. It follows from the above discussion that \mathbf{K} is a partition of P^* since \mathbf{V} and \mathbf{U} are partitions of $((P \setminus H)^*H)^*$ and $(P \setminus H)^*$, respectively. Moreover, every $K \in \mathbf{K}$ belongs to $SD(\mathcal{C})$: $K = VU$ with $V \in \mathbf{V}$ and $U \in \mathbf{U}$, and one may verify that this is a *unambiguous* concatenation. It remains to show that \mathbf{K} is s -safe. Let $K \in \mathbf{K}$ and $w, w' \in K$. We show that $s\alpha(w) = s\alpha(w')$. By definition, we have $K = VU$ with $V \in \mathbf{V}$ and $U \in \mathbf{U}$. Therefore, $w = vu$ and $w' = v'u'$ with $u, u' \in U$ and $v, v' \in V$. Since U and V are both 1_M -safe by definition, we have $\alpha(u) = \alpha(u')$ and $\alpha(v) = \alpha(v')$. It follows that $s\alpha(w) = s\alpha(w')$, which concludes the proof of Proposition 9.

B Appendix to Section 4

This is devoted to the proof of Theorem 20: the characterization of $SF(\mathcal{C})$ -optimal imprints which is generic to all finite quotient-closed Boolean algebras \mathcal{C} . We fix \mathcal{C} for the presentation. First, we recall the theorem and then concentrate on its proof.

B.1 Characterization

In view of Proposition 19, deciding $SF(\mathcal{C})$ -covering amounts to computing $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from an input nice multiplicative rating map ρ . Our algorithm actually computes slightly more information.

Since \mathcal{C} is a finite quotient-closed Boolean algebra, we may consider the equivalence $\sim_{\mathcal{C}}$ over A^* . In particular, the set $A^*/\sim_{\mathcal{C}}$ of $\sim_{\mathcal{C}}$ -classes is a finite monoid (we write “ \cdot ” for its multiplication) and the map $w \mapsto [w]_{\mathcal{C}}$ is a morphism. Given a rating map $\rho : 2^{A^*} \rightarrow R$ we define:

$$\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho] = \{(C, r) \in (A^*/\sim_{\mathcal{C}}) \times R \mid r \in \mathcal{I}_{SF(\mathcal{C})}[C, \rho]\}$$

Observe that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ captures more information than $\mathcal{I}_{SF(\mathcal{C})}[\rho]$. Indeed, it encodes all sets $\mathcal{I}_{SF(\mathcal{C})}[C, \rho]$ for $C \in A^*/\sim_{\mathcal{C}}$ and by Fact 18, $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ is the union of all these sets.

Our main result is a least fixpoint procedure for computing $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ from a nice multiplicative rating map ρ . It is based on a generic characterization theorem which we first present. Given an arbitrary nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$ and a set $S \subseteq (A^*/\sim_{\mathcal{C}}) \times R$, we say that S is $SF(\mathcal{C})$ -saturated for ρ when the following properties are satisfied:

1. **Trivial elements.** For every $w \in A^*$, we have $([w]_{\mathcal{C}}, \rho(w)) \in S$.
2. **Downset.** For every $(C, r) \in S$ and $q \in R$, if $q \leq r$, then $(C, q) \in S$.
3. **Multiplication.** For every $(C, q), (D, r) \in S$, we have $(C \cdot D, qr) \in S$.
4. **$SF(\mathcal{C})$ -closure.** For all $(E, r) \in S$, if $E \in A^*/\sim_{\mathcal{C}}$ is idempotent, then $(E, r^\omega + r^{\omega+1}) \in S$.

► **Theorem 20** ($SF(\mathcal{C})$ -optimal imprints (\mathcal{C} finite)). *Let $\rho : 2^{A^*} \rightarrow R$ be a nice multiplicative rating map. Then, $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ is the least $SF(\mathcal{C})$ -saturated subset of $(A^*/\sim_{\mathcal{C}}) \times R$ for ρ .*

Given a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$ as input, it is clear the one may compute the least $SF(\mathcal{C})$ -saturated subset of $(A^*/\sim_{\mathcal{C}}) \times R$ with a least fixpoint procedure. Hence, Theorem 20 provides an algorithm for computing $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$. As we explained above, we may then compute $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from this set. Together with Proposition 19, this yields Theorem 2 as a corollary: $SF(\mathcal{C})$ -covering is decidable when \mathcal{C} is a finite quotient-closed Boolean algebra.

B.2 Proof

We turn to the proof of Theorem 20. Let us fix a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$ for the argument. We prove that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ is the least $SF(\mathcal{C})$ -saturated subset of $(A^*/\sim_{\mathcal{C}}) \times R$ (for ρ). The argument involves two directions which correspond respectively to soundness and completeness of the least fixpoint procedure which computes $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ from ρ .

- **Soundness:** We prove that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ is $SF(\mathcal{C})$ -saturated.
- **Completeness:** We prove that every $SF(\mathcal{C})$ -saturated set is included in $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$.

Soundness

First, we prove that the set $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho] \subseteq (A^*/\sim_{\mathcal{C}}) \times R$ itself is $SF(\mathcal{C})$ -saturated. The argument is based on Lemma 34.

► **Remark 36.** We do not need the hypothesis that ρ is nice for this direction of the proof.

That $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ contains the trivial elements and is closed under downset and multiplication is actually a generic property of optimal imprints: this hold as soon as the investigated class is a quotient-closed Boolean algebra (see Lemma 9.5 in [22]). This is the case for $SF(\mathcal{C})$ since \mathcal{C} is a quotient-closed Boolean algebra by hypothesis. Hence, we concentrate on proving that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ satisfies $SF(\mathcal{C})$ -closure. Consider $(E, r) \in \mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ such that $E \in A^*/\sim_{\mathcal{C}}$ is idempotent. We show that $(E, r^\omega + r^{\omega+1}) \in \mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$. By definition, this corresponds to the following property:

$$r^\omega + r^{\omega+1} \in \mathcal{I}_{SF(\mathcal{C})}[E, \rho].$$

This amounts to proving that for every $SF(\mathcal{C})$ -cover \mathbf{K} of E , we have $r^\omega + r^{\omega+1} \in \mathcal{I}[\rho](\mathbf{K})$. We fix \mathbf{K} for the proof: we have to exhibit $K \in \mathbf{K}$ such that $r^\omega + r^{\omega+1} \leq \rho(K)$. We start with a few definitions that we require to describe K . Since \mathbf{K} is finite and $SF(\mathcal{C})$ is quotient-closed Boolean algebra, we have the following fact (see Lemma 17 in [23] for a proof).

► **Fact 37.** *There exists a finite quotient-closed Boolean algebra \mathcal{D} such that $\mathcal{D} \subseteq SF(\mathcal{C})$ and every $K \in \mathbf{K}$ belongs to \mathcal{D} .*

We fix \mathcal{D} as the finite quotient-closed Boolean algebra given by Fact 37. Recall that $\sim_{\mathcal{D}}$ denotes the associated canonical equivalence defined on A^* . Since \mathcal{D} is closed under quotients we know that $\sim_{\mathcal{D}}$ is a congruence for word concatenation. Additionally, we have the following more involved property which is a corollary of Lemma 34.

► **Lemma 38.** *There exists a natural number $k \in \mathbb{N}$ such that for every $\ell \geq k$ and $u \in E$, we have $u^\ell \sim_{\mathcal{D}} u^\ell u$.*

Proof. Lemma 34 yields that for every language $L \in \mathcal{D} \subseteq SF(\mathcal{C})$, there exists $k_L \in \mathbb{N}$ such that for every $\ell \geq k_L$ and every $u \in A^*$ satisfying $uu \sim_{\mathcal{C}} u$, we have $u^\ell \in L$ if and only if $u^\ell u \in L$. We choose k as the maximum of all numbers k_L for $L \in \mathcal{D}$ (recall that \mathcal{D} is finite). It remains to show that the lemma holds for this k . Consider $\ell \geq k$ and $u \in E$. Since E is an

121:26 On all things star-free

idempotent $\sim_{\mathcal{C}}$ -class, we have $uu \sim_{\mathcal{C}} u$. Hence, by choice of k , it is immediate that $u^\ell \in L$ if and only if $u^\ell u \in L$ for every $L \in \mathcal{D}$. By definition, this exactly says that $u^\ell \sim_{\mathcal{D}} u^\ell u$. ◀

We may now come back to the main argument. We write \mathbf{H} for the set of all $\sim_{\mathcal{D}}$ -classes which intersect E . Clearly, \mathbf{H} is a \mathcal{D} -cover of E and therefore an $SF(\mathcal{C})$ -cover of E since $\mathcal{D} \subseteq SF(\mathcal{C})$ by definition in Fact 37. Hence, since $(E, r) \in \mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ by hypothesis (which means that $r \in \mathcal{I}_{SF(\mathcal{C})}[E, \rho]$), we obtain that $r \in \mathcal{I}[\rho](\mathbf{H})$. This yields $H \in \mathbf{H}$ such that $r \leq \rho(H)$. Consider the natural number $k \in \mathbb{N}$ given by Lemma 38 and the idempotent power ω of R for multiplication. We define,

$$G = H^{k\omega} \cup H^{k\omega+1}$$

The argument is now based on the following lemma which exhibits the desired language $K \in \mathbf{K}$ such that $r^\omega + r^{\omega+1} \leq \rho(K)$.

► **Lemma 39.** *The exists $K \in \mathbf{K}$ such that $G \subseteq K$.*

Before we prove Lemma 39, let us use it to conclude the argument. By definition of G and since $r \leq \rho(H)$, we get,

$$r^\omega + r^{\omega+1} \leq \rho(H^{k\omega} \cup H^{k\omega+1}) = \rho(G)$$

Furthermore, since the language $K \in \mathbf{K}$ given by Lemma 39 satisfies $G \subseteq K$, we obtain that $r^\omega + r^{\omega+1} \leq \rho(K)$ as desired. This yields $r^\omega + r^{\omega+1} \in \mathcal{I}[\rho](\mathbf{K})$ finishing the proof. We now prove Lemma 39.

Proof of Lemma 39. Since $H \in \mathbf{H}$, we know that H is a $\sim_{\mathcal{D}}$ -class intersecting E . Let $u \in H \cap E$ and $v = u^{k\omega}$. Clearly, $v \in E^{k\omega}$ and since E is an idempotent of $A^*/\sim_{\mathcal{C}}$, it is immediate that $E^{k\omega} \subseteq E$ which yields $v \in E$. By hypothesis, \mathbf{K} is a cover of E . Thus, we have $K \in \mathbf{K}$ such that $v \in E$. We show that $G \subseteq K$ which concludes the proof.

Consider $w \in G$. We have to prove that $w \in K$. We show that $w \sim_{\mathcal{D}} v$. Since $v \in K$ by definition of K and $K \in \mathcal{D}$ (see Fact 37), this implies that $w \in K$. By definition of G , either $w \in H^{k\omega}$ or $w \in H^{k\omega+1}$. We treat the two cases separately. Assume first that $w \in H^{k\omega}$. Recall that $u \in H$ and H is a $\sim_{\mathcal{D}}$ -class. Thus, since $\sim_{\mathcal{D}}$ is a congruence, it follows that $w \sim_{\mathcal{D}} u^{k\omega} = v$ which concludes the proof. Assume now that $w \in H^{k\omega+1}$. Again, since $\sim_{\mathcal{D}}$ is a congruence, we get that $w \sim_{\mathcal{D}} u^{k\omega+1}$. Moreover, we have $u \in E$. Therefore, Lemma 38 yields that $u^{k\omega+1} \sim_{\mathcal{D}} u^{k\omega} = v$. Transitivity then yields $w \sim_{\mathcal{D}} v$ which concludes the proof. ◀

Completeness

We turn to the most interesting direction in Theorem 20. We show that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$ is included in every $SF(\mathcal{C})$ -saturated subset of $(A^*/\sim_{\mathcal{C}}) \times R$ for ρ . Consequently, we fix $S \subseteq (A^*/\sim_{\mathcal{C}}) \times R$ which is $SF(\mathcal{C})$ -saturated for the proof. As expected, the argument involves building a particular $SF(\mathcal{C})$ -cover.

We start with some terminology that we need to present the argument. It simplifies the presentation to use “ $(A^*/\sim_{\mathcal{C}}) \times R$ ” as the rating set of a new rating map γ that we build from ρ . However, since $A^*/\sim_{\mathcal{C}}$ is not a semiring, we need to slightly modify this set. Since $A^*/\sim_{\mathcal{C}}$ is a finite monoid, $2^{A^*/\sim_{\mathcal{C}}}$ is clearly an idempotent semiring: the addition is union and the multiplication is lifted from the one of $A^*/\sim_{\mathcal{C}}$. Hence, $Q = 2^{A^*/\sim_{\mathcal{C}}} \times R$ is an idempotent semiring for the componentwise addition and multiplication. Consider the map $\gamma : 2^{A^*} \rightarrow Q$ defined by $\gamma(K) = (\{[w]_{\mathcal{C}} \mid w \in K\}, \rho(K))$ for every $K \subseteq A^*$. It is straightforward to verify

that γ is a nice multiplicative rating map since ρ is one. Moreover, we reformulate our $SF(\mathcal{C})$ -saturated set $S \subseteq (A^*/\sim_{\mathcal{C}}) \times R$ as the following subset T of Q :

$$T = \{(\{C\}, r) \mid (C, r) \in S\} \subseteq Q = 2^{A^*/\sim_{\mathcal{C}}} \times R$$

Since S is $SF(\mathcal{C})$ -saturated for ρ , we know that $([w]_{\mathcal{C}}, \rho(w)) \in S$ for every $w \in A^*$ (this is a trivial element) and S is closed under multiplication. By definition of T , this implies that $\tau(w) \in T$ for every $w \in A^*$ (in particular $1_Q = \tau(\varepsilon) \in T$) and T is closed under multiplication. From now on, we assume that these two properties of T are understood.

Finally, for every finite set of languages \mathbf{H} , we associate two subsets of Q . The definitions are as follows:

- $Q_{\mathbf{H}}^+ \subseteq Q$ is the least subset of Q closed under addition and multiplication such that $\rho(H) \in Q_{\mathbf{H}}^+$ for every $H \in \mathbf{H}$.
- $Q_{\mathbf{H}}^* \subseteq Q$ is the least subset of Q closed under addition and multiplication such that $1_Q \in Q_{\mathbf{H}}^*$ and $\rho(H) \in Q_{\mathbf{H}}^*$ for every $H \in \mathbf{H}$.

We are ready to present the completeness argument. It is based on the following statement which generalizes Proposition 9. We prove it by induction. Recall that given a language $P \subseteq A^*$, a $SD(\mathcal{C})$ -partition of P is a finite partition of P into languages of $SD(\mathcal{C}) = SF(\mathcal{C})$.

► **Proposition 40.** *Let $P \subseteq A^+$ be a prefix code with bounded synchronization delay and \mathbf{H} a $SD(\mathcal{C})$ -partition of P such that $\gamma(H) \in T$ for every $H \in \mathbf{H}$. Then, for every $t \in T$, there exists a $SD(\mathcal{C})$ -partition \mathbf{K} of P^* satisfying the following property,*

$$\text{for every } K \in \mathbf{K}, \gamma(K) \in Q_{\mathbf{H}}^* \text{ and } t \cdot \gamma(K) \in T \quad (5)$$

Before we prove Proposition 40, let us first apply it to prove the completeness direction in Theorem 20. We start by constructing a $SD(\mathcal{C})$ -partition \mathbf{K} of A^* with the proposition.

Observe that $A \subseteq A^+$ is a prefix code with bounded synchronization delay. Moreover, $\mathbf{H} = \{\{a\} \mid a \in A\}$ is a $SD(\mathcal{C})$ -partition of A and we have $\gamma(a) \in T$ for every $a \in A$. Therefore, we may apply Proposition 40 in the case when $P = A$ and $t = 1_Q \in T$. This yields a $SD(\mathcal{C})$ -partition \mathbf{K} of A^* satisfying (5). We use it to prove that $\mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho] \subseteq S$.

Let $(C, r) \in \mathcal{P}_{SF(\mathcal{C})}^{\mathcal{C}}[\rho]$, i.e. $r \in \mathcal{I}_{SF(\mathcal{C})}[C, \rho]$. We define $\mathbf{K}_C \subseteq \mathbf{K}$ as the set containing every language $K \in \mathbf{K}$ such that $K \cap C \neq \emptyset$. Clearly, \mathbf{K}_C is a $SD(\mathcal{C})$ -cover of C since \mathbf{K} is a $SD(\mathcal{C})$ -cover of A^* . Moreover, it is a $SF(\mathcal{C})$ -cover of C since $SF(\mathcal{C}) = SD(\mathcal{C})$ by Theorem 7. It follows that $\mathcal{I}_{SF(\mathcal{C})}[C, \rho] \subseteq \mathcal{I}[\rho](\mathbf{K}_C)$ and we obtain $r \in \mathcal{I}[\rho](\mathbf{K}_C)$. This yields a language $K \in \mathbf{K}_C$ such that $r \leq \rho(K)$. Moreover, $K \in \mathbf{K}$ and (5) yields that $\gamma(K) \in T$ (recall that we chose $t = 1_Q$). By definition of T , it follows that $\gamma(K) = (\{D\}, p)$ for some $(D, p) \in S$. By definition of γ , $p = \rho(K)$ and since $K \cap C \neq \emptyset$, we have $D = C$. Hence, $(C, \rho(K)) \in S$ and since $r \leq \rho(K)$, closure under downset for S (recall that S is $SF(\mathcal{C})$ -saturated) yields $(C, r) \in S$, concluding the proof.

It remains to prove Proposition 40. We let $P \subseteq A^+$ be a prefix code with bounded synchronization delay and \mathbf{H} a $SD(\mathcal{C})$ -partition of P such that $\gamma(H) \in T$ for every $H \in \mathbf{H}$. Finally, we fix $t \in T$. We need to build a $SD(\mathcal{C})$ -partition \mathbf{K} of P^* satisfying (5). We proceed by induction on the three following parameters, listed by order of importance:

1. The size of the set $Q_{\mathbf{H}}^+ \subseteq Q$,
2. The size of \mathbf{H} ,
3. The size of the set $t \cdot Q_{\mathbf{H}}^* \subseteq Q$.

We distinguish two main cases depending on the following property. We say that t is **H-stable** when the following holds,

$$\text{for every } H \in \mathbf{H}, \quad t \cdot Q_{\mathbf{H}}^* = t \cdot Q_{\mathbf{H}}^* \cdot \gamma(H) \quad (6)$$

We first consider the case when t is **H-stable**. This is the base case: we construct **K** directly. Then, we handle the converse case using induction on our three parameters.

Base case: t is **H-stable**

In this case, we define **K** directly: **K** contains all languages $P^* \cap C$ for $C \in A^*/\sim_{\mathcal{C}}$ which are **nonempty**. Clearly, this is a partition of P^* . Moreover, it only contains languages in $SD(\mathcal{C})$. Indeed, we have $P \in SD(\mathcal{C})$: it is the disjoint union of all languages in the $SD(\mathcal{C})$ -partition **H** of P . Hence, $P^* \in SD(\mathcal{C})$ since P is a prefix code with bounded synchronization delay. Finally, it follows that $P^* \cap C \in SD(\mathcal{C})$ for every $C \in A^*/\sim_{\mathcal{C}}$ since $C \in \mathcal{C}$. We have to prove **K** satisfies (5): given $K \in \mathbf{K}$, we show that $\gamma(K) \in Q_{\mathbf{H}}^*$ and $t \cdot \gamma(K) \in T$.

We begin with an important definition. A **H-product** is a language of the form $H_1 \cdots H_n$ with $n \in \mathbb{N}$ and $H_1, \dots, H_n \in \mathbf{H}$ (this includes the case $n = 0$: $\{\varepsilon\}$ is a **H-product**). Clearly, $\gamma(V) \in Q_{\mathbf{H}}^*$, for every **H-product** V . Moreover, since $\gamma(H) \in T$ for every $H \in \mathbf{H}$ and T is closed under multiplication and contains 1_Q , we also have $\gamma(V) \in T$ for every **H-product** V . Using this observation, we prove two simple properties of **H-products**.

► **Fact 41.** *Every **H-product** V is nonempty and there exists a unique $C \in A^*/\sim_{\mathcal{C}}$, denoted by $[V]_{\mathcal{C}}$ such that $V \subseteq C$.*

Proof. Since $\gamma(V) \in T$, we get by definition of T that $\gamma(V) = (\{C\}, r)$ for some $C \in A^*/\sim_{\mathcal{C}}$ and $r \in R$. By definition of γ , we have $\{C\} = \{[w]_{\mathcal{C}} \mid w \in V\}$. Thus, V is nonempty and $V \subseteq C$. ◀

We turn to an important property of **H-products**: they are the building blocks of the languages within our $SD(\mathcal{C})$ -partition **K** of P^* .

► **Fact 42.** *For every $C \in A^*/\sim_{\mathcal{C}}$, $P^* \cap C$ is the (possibly infinite or empty) union of all **H-products** V such that $[V]_{\mathcal{C}} = C$.*

Proof. By definition, **H** is a partition of P . Hence, $P^* \cap C$ is the (infinite) union of all languages $H_1 \cdots H_n \cap C$ for $n \in \mathbb{N}$ and $H_1, \dots, H_n \in \mathbf{H}$. Consider such a language. Clearly, $H_1 \cdots H_n$ is a **H-product**. Therefore, either $[H_1 \cdots H_n]_{\mathcal{C}} = C$ and $H_1 \cdots H_n \cap C = H_1 \cdots H_n$ or $[H_1 \cdots H_n]_{\mathcal{C}} \neq C$ and $H_1 \cdots H_n \cap C = \emptyset$. The fact follows. ◀

We may start the proof that **K** satisfies (5). First, we show that $\gamma(K) \in Q_{\mathbf{H}}^*$ for every $K \in \mathbf{K}$. By definition, K is a nonempty language of the form $P^* \cap C$ with $C \in A^*/\sim_{\mathcal{C}}$. Hence, Fact 42 yields that K is a (possibly infinite) union of **H-products**. Since γ is nice by definition, it follows that there exists finitely many **H-products** V_1, \dots, V_ℓ such that $\gamma(K) = \gamma(V_1) + \cdots + \gamma(V_\ell)$. We know that $\gamma(V_i) \in Q_{\mathbf{H}}^*$ for every $i \leq \ell$. Therefore, $\gamma(K) \in Q_{\mathbf{H}}^*$ since $Q_{\mathbf{H}}^*$ is closed under addition by definition.

It remains to show that $t\gamma(K) \in T$ for every $K \in \mathbf{K}$. This is where we need the hypothesis that t is **H-stable**. We use it to prove the following preliminary lemma.

► **Lemma 43.** *Let V be an **H-product** such that $\gamma(V)$ is a multiplicative idempotent of Q . Then, for every $q \in Q_{\mathbf{H}}^*$, we have $tq\gamma(V) = tq$.*

Proof. We first use the hypothesis that t is \mathbf{H} -stable to prove the following preliminary result which holds regardless of whether $\gamma(V)$ is idempotent:

$$\text{there exists } x \in Q_{\mathbf{H}}^* \text{ such that } tx\gamma(V) = tq \tag{7}$$

Since V is a \mathbf{H} -product, there exists $n \in \mathbb{N}$ and $H_1, \dots, H_n \in \mathbf{H}$ such that $V = H_1 \cdots H_n$. We proceed by induction on n . If $n = 0$, then $V = \{\varepsilon\}$ and $\gamma(V) = 1_Q$, it suffices to choose $x = q \in Q_{\mathbf{H}}^*$. Otherwise $V = HV'$ for $H \in \mathbf{H}$ and V' a \mathbf{H} -product of smaller length. Using induction, we get $y \in Q_{\mathbf{H}}^*$ such that $ty\gamma(V') = tq$. Moreover, since t is \mathbf{H} -stable, we get from (6) that there exists $x \in Q_{\mathbf{H}}^*$ such that $ty = tx\gamma(H)$. It follows that $tx\gamma(V) = tx\gamma(H)\gamma(V') = ty\gamma(V') = tq$. This concludes the proof of (7). It remains to prove the lemma.

We have $tx\gamma(V) = tq$. Hence, if $\gamma(V)$ is a multiplicative idempotent of Q , we get $tq\gamma(V) = tx\gamma(V)\gamma(V) = tx\gamma(V) = tq$. This completes the proof. ◀

We are ready to prove that $t\gamma(K) \in T$ for every $K \in \mathbf{K}$. We fix $K \in \mathbf{K}$ for the proof. By definition, K is non empty and $K = P^* \cap C$ for $C \in A^*/\sim_{\mathcal{C}}$. We first treat the case when C is an idempotent E of $A^*/\sim_{\mathcal{C}}$. Then we use it to treat the general case.

Idempotent case. By Fact 42, $K = P^* \cap E$ is the union of all \mathbf{H} -products V such $[V]_{\mathcal{C}} = E$. This union is non-empty since K is nonempty. Since γ is nice, this yields finitely many \mathbf{H} -products V_1, \dots, V_ℓ such that $\gamma(K) = \gamma(V_1) + \dots + \gamma(V_\ell)$ and $[V_i]_{\mathcal{C}} = E$ for every $i \leq \ell$.

Consider $i \leq \ell$. Moreover, $[V_i]_{\mathcal{C}} = E$ means that $V_i \neq \emptyset$ and $V_i \subseteq E$ by Fact 41. Therefore, $\gamma(V_i) = (\{E\}, \rho(V_i))$ by definition of γ . Moreover, we know that $\gamma(V_i) \in T$ which yields $(E, \rho(V_i)) \in S$ by definition of T . Since E is idempotent and S is $SF(\mathcal{C})$ -saturated, we obtain from $SD(\mathcal{C})$ -closure that $(E, (\rho(V_i))^\omega + (\rho(V_i))^{\omega+1}) \in S$ (here, we use “ ω ” to denote a multiplicative idempotent power for both R and Q). Since this holds for every $i \leq \ell$ and E is idempotent, it then follows from closure under multiplication for S that,

$$(E, \prod_{1 \leq i \leq \ell} ((\rho(V_i))^\omega + (\rho(V_i))^{\omega+1})) \in S$$

For $i \leq \ell$, let $r_i = (\rho(V_1))^\omega \cdots (\rho(V_{i-1}))^\omega (\rho(V_i))^{\omega+1} (\rho(V_{i+1}))^\omega \cdots (\rho(V_\ell))^\omega \in R$. One may verify that $r_1 + \dots + r_\ell \leq \prod_{1 \leq i \leq \ell} ((\rho(V_i))^\omega + (\rho(V_i))^{\omega+1})$ by distributing the multiplication in the right side of this inequality. Since S is closed under downset (it is $SF(\mathcal{C})$ -saturated), this yields $(E, r_1 + \dots + r_\ell) \in S$. By definition of T , we get $(\{E\}, r_1 + \dots + r_\ell) \in T$.

For $i \leq \ell$, let $q_i = (\gamma(V_1))^\omega \cdots (\gamma(V_{i-1}))^\omega (\gamma(V_i))^{\omega+1} (\gamma(V_{i+1}))^\omega \cdots (\gamma(V_\ell))^\omega \in Q$. Since $\{E\}$ is idempotent for both addition and multiplication in $2^{A^*/\sim_{\mathcal{C}}}$, and $\gamma(V_i) = (\{E\}, \rho(V_i))$, one may verify that $q_1 + \dots + q_\ell = (\{E\}, r_1 + \dots + r_\ell) \in T$. Since $t \in T$ by definition and T is closed under multiplication, we obtain $t(q_1 + \dots + q_\ell) \in T$. Moreover, one may verify from Lemma 43 that for every $i \leq \ell$, we have $tq_i = t\gamma(V_i)$. Altogether, this yields that $t\gamma(K) = t(q_1 + \dots + q_\ell) \in T$ which concludes the idempotent case.

General case. It remains to handle the case when $K = P \cap C^*$ for an arbitrary $C \in A^*/\sim_{\mathcal{C}}$. By Fact 42, K is the union of all \mathbf{H} -products V such that $[V]_{\mathcal{C}} = C$. Since K is non empty, there exists a least one such V . We fix it for the proof. Since (Q, \cdot) is a finite monoid, there exists a number $p \geq 1$ such that $\gamma(V^p)$ is a multiplicative idempotent of Q . Since $[V]_{\mathcal{C}} = C$, we have $\gamma(V) = (\{C\}, \rho(V))$. Hence, $\gamma(V^p) = (\{C\}^p, \rho(V^p))$ and the multiplication of p copies of C with “ \cdot ” is an idempotent $E \in A^*/\sim_{\mathcal{C}}$ (in particular, we have $C^p \subseteq E$).

We know that $P^* \cap E \neq \emptyset$ (it includes all words in V^p). Hence, since we already handled the idempotent case, we know that,

$$(\{E\}, \rho(P^* \cap E)) = \tau(P^* \cap E) \in T$$

121:30 On all things star-free

This yields $(E, \rho(P^* \cap E)) \in S$ by definition of T . We have $P^*V^{p-1} \subseteq P^*$ since $V \subseteq P^*$ (V is the concatenation of languages in \mathbf{H} which is a partition of P). Moreover, $CV^{p-1} \subseteq E$ since $V \subseteq C$ and $C^p \subseteq E$. Altogether, we obtain $(P^* \cap C)V^{p-1} \subseteq P^* \cap E$. Therefore, $\rho((P^* \cap C)V^{p-1}) \subseteq \rho(P^* \cap E)$ and since S is closed under downset (it is $SF(\mathcal{C})$ -saturated), this yields $(E, \rho((P^* \cap C)V^{p-1})) \in S$. By definition of T , this implies that,

$$\gamma((P^* \cap C)V^{p-1}) = (\{E\}, \rho((P^* \cap C)V^{p-1})) \in T$$

Moreover, we have $t, \gamma(V) \in T$. Since T is closed under multiplication, we get that $t\gamma(K)\gamma(V^p) = t\gamma(P^* \cap C)\gamma(V^p) \in T$. We already established that $\gamma(K) \in Q_{\mathbf{H}}^*$. Moreover, $\gamma(V^p)$ is idempotent by definition. Hence, Lemma 43 yields that $t\gamma(K)\gamma(V^p) = t\gamma(K)$. Altogether, we obtain $t\gamma(K) \in T$, concluding the proof.

Inductive step: t is not \mathbf{H} -stable

The hypothesis that t is not \mathbf{H} -stable yields some language $H \in \mathbf{H}$ such that the following **strict** inclusion holds:

$$t \cdot Q_{\mathbf{H}}^* \cdot \gamma(H) \subsetneq t \cdot Q_{\mathbf{H}}^* \tag{8}$$

We fix this language $H \in \mathbf{H}$ for the remainder of the proof. Using induction on our second parameter in Proposition 40, we prove the following lemma.

► **Lemma 44.** *There exists a $SD(\mathcal{C})$ -partition \mathbf{U} of $(P \setminus H)^*$ such that for every $U \in \mathbf{U}$, we have $\gamma(U) \in Q_{\mathbf{H}}^*$ and $\gamma(U) \in T$.*

Proof. Fact 28 implies that $P \setminus H$ remains a prefix code with bounded synchronization delay since P was one. We want to apply induction in Proposition 40 for the case when P has been replaced by $P \setminus H$. This requires defining a $SD(\mathcal{C})$ -partition \mathbf{G} of $P \setminus H$ and verifying that our induction parameters have decreased. We let $\mathbf{G} = \mathbf{H} \setminus \{H\}$ which is a $SD(\mathcal{C})$ -partition of $P \setminus H$ since \mathbf{H} was a $SD(\mathcal{C})$ -partition of P . Moreover, $\tau(\mathbf{G}) \in T$ for every $G \in \mathbf{G}$ by hypothesis on \mathbf{H} . Finally, it is immediate that $Q_{\mathbf{G}}^+ \subseteq Q_{\mathbf{H}}^+$ (our first induction has not increased) and $\mathbf{G} \subsetneq \mathbf{H}$ (our second parameter has decreased). Hence, we may apply Proposition 40 in the case when P, \mathbf{H} and $t \in T$ have been replaced by $P \setminus H, \mathbf{G}$ and $1_Q \in T$. This yields a $SD(\mathcal{C})$ -partition \mathbf{U} of $(P \setminus H)^*$ such that for every $U \in \mathbf{U}$, we have $\gamma(U) \in Q_{\mathbf{G}}^* \subseteq Q_{\mathbf{H}}^*$ and $\gamma(U) \in T$. ◀

We fix the $SD(\mathcal{C})$ -partition \mathbf{U} of $(P \setminus H)^*$ given by Lemma 44 for the remainder of the proof. We distinguish two sub-cases.

Since $H \in \mathbf{H}$, one may verify from the definitions of $Q_{\mathbf{H}}^+$ and $Q_{\mathbf{H}}^*$ that $Q_{\mathbf{H}}^* \cdot \gamma(H) \subseteq Q_{\mathbf{H}}^+$. We consider two sub-cases depending on whether this inclusion is strict or not.

Sub-case 1: we have the equality $Q_{\mathbf{H}}^* \cdot \gamma(H) = Q_{\mathbf{H}}^+$

Recall that we have to exhibit a $SD(\mathcal{C})$ -partition \mathbf{K} of P^* which satisfies (5). In this case, the argument is based on the following lemma which is proved using our hypotheses and induction on our third parameter (i.e. the size of $t \cdot Q_{\mathbf{H}}^*$).

► **Lemma 45.** *For every $U \in \mathbf{U}$, there exists a $SD(\mathcal{C})$ -partition \mathbf{W}_U of P^* such that for every $W \in \mathbf{W}_U$, $\gamma(W) \in Q_{\mathbf{H}}^*$ and $t\gamma(UHW) \in T$.*

Proof. We fix $U \in \mathbf{U}$ for the proof. By definition of \mathbf{U} in Lemma 44, $\gamma(U) \in Q_{\mathbf{H}}^*$ and $\gamma(U) \in T$. Hence, since $t, \gamma(H) \in T$ by hypothesis and T is closed under multiplication, we have $t\gamma(U)\gamma(H) \in T$. Moreover, it is clear that $t\gamma(U)\gamma(H) \cdot Q_{\mathbf{H}}^* \subseteq t \cdot Q_{\mathbf{H}}^* \cdot \gamma(H) \cdot Q_{\mathbf{H}}^* \subseteq t \cdot Q_{\mathbf{H}}^+$. Combined with our hypothesis in Sub-case 1 (i.e. $Q_{\mathbf{H}}^* \cdot \gamma(H) = Q_{\mathbf{H}}^+$), this yields $t\gamma(U)\gamma(H) \cdot Q_{\mathbf{H}}^* \subseteq t \cdot Q_{\mathbf{H}}^+$. We may then use (8) ($t \cdot Q_{\mathbf{H}}^* \cdot \gamma(H) \subsetneq t \cdot Q_{\mathbf{H}}^+$) to get the **strict** inclusion $t\gamma(U)\gamma(H) \cdot Q_{\mathbf{H}}^* \subseteq t \cdot Q_{\mathbf{H}}^+$. Consequently, we may use induction on our third parameter (i.e. the size of $t \cdot Q_{\mathbf{H}}^*$) to apply Proposition 40 in the case when $t \in T$ has been replaced by $t\gamma(U)\gamma(H) \in T$. Note that here, our first two parameters have not increased (they only depend on \mathbf{H} which remains unchanged). This yields the desired $SD(\mathcal{C})$ -partition \mathbf{W}_U of P^* . ◀

It remains to use Lemma 45 to conclude the proof of Sub-case 1. We build our $SD(\mathcal{C})$ -partition \mathbf{K} of P^* as follows,

$$\mathbf{K} = \mathbf{U} \cup \bigcup_{U \in \mathbf{U}} \{UHW \mid W \in \mathbf{W}_U\}$$

It remains to show that \mathbf{K} is a $SD(\mathcal{C})$ -partition of P^* which satisfies (5). First, observe that the languages $K \in \mathbf{K}$ belong to $SD(\mathcal{C})$. This is immediate by definition of \mathbf{U} is $K \in \mathbf{U}$. Otherwise, $K = UHW$ with $U \in \mathbf{U}$ and $W \in \mathbf{W}_U$ and $U, H, W \in SD(\mathcal{C})$. Moreover, one may verify that the concatenation UHW is *unambiguous* since P is a prefix code, $U \subseteq (P \setminus H)^*$ and $W \subseteq H^*$. Altogether, it follows that $K \in SD(\mathcal{C})$.

That \mathbf{K} is a partition of P^* is also simple to verify from the definition since P is a prefix code, $H \subseteq P$ and \mathbf{U}, \mathbf{W}_U are partitions of $(P \setminus H)^*$ and P^* respectively. Every word $w \in P^*$ admits a *unique* decomposition $w = w_1 \cdots w_n$ with $w_1, \dots, w_n \in P$. We partition P^* by looking at the leftmost factor belonging to H (when it exists).

It remains to prove that (5) holds. Consider $K \in \mathbf{K}$, we show that $\gamma(K) \in Q_{\mathbf{H}}^*$ and $t\gamma(K) \in T$. If $K \in \mathbf{U}$, this is immediate by definition of \mathbf{U} in Lemma 44. Otherwise, $K = UHW$ with $U \in \mathbf{U}$ and $W \in \mathbf{W}_U$. By definition of \mathbf{U} and \mathbf{W}_U , we have $\gamma(U), \gamma(W) \in Q_{\mathbf{H}}^*$. Hence, $\gamma(UHW) \in Q_{\mathbf{H}}^*$. Moreover, we have $t\gamma(UHW) \in T$ by definition of \mathbf{W}_U in Lemma 45. This concludes the first sub-case.

Sub-case 2: we have the strict inclusion $Q_{\mathbf{H}}^* \cdot \gamma(H) \subsetneq Q_{\mathbf{H}}^+$

Recall that our objective is to construct a $SD(\mathcal{C})$ -partition \mathbf{K} of P^* which satisfies (5). We begin by giving a brief overview of the construction. Consider a word $w \in P^*$. Since P is a prefix code, w admits a unique decomposition as a concatenation of factors in P . We may look at the rightmost factor in $H \subseteq P$ to uniquely decompose w in two parts (each of them possibly empty): a prefix in $((P \setminus H)^*H)^*$ and a suffix in $(P \setminus H)^*$. We use induction to construct $SD(\mathcal{C})$ -partitions of the sets of possible prefixes and suffixes. Then, we combine them to construct a $SD(\mathcal{C})$ -partition of the whole set P^* . Actually, we already constructed a suitable $SD(\mathcal{C})$ -partition of the possible suffixes in $(P \setminus H)^*$: \mathbf{U} (see Lemma 44). Hence, it remains to partition the prefixes. We do so this in the following lemma which is proved using the hypothesis of Sub-case 2 and induction on our first parameter.

► **Lemma 46.** *There exists a $SD(\mathcal{C})$ -partition \mathbf{V} of $((P \setminus H)^*H)^*$ such that for every $V \in \mathbf{V}$, $\gamma(V) \in Q_{\mathbf{H}}^*$ and $\gamma(V) \in T$.*

Proof. Let $L = (P \setminus H)^*H$. Fact 29 implies that L remains a prefix code with bounded synchronization delay since P was one. We want to apply induction in Proposition 40

for the case when P has been replaced by L . Doing so requires building an appropriate $SD(\mathcal{C})$ -partition of L and proving that one of our induction parameters has decreased.

Let $\mathbf{F} = \{UH \mid U \in \mathbf{U}\}$. Since \mathbf{U} is a partition of $(P \setminus H)^*$ and P is a prefix code, one may verify that \mathbf{F} is a partition of $L = (P \setminus H)^*H$. Moreover, it is a $SD(\mathcal{C})$ -partition of L : every $F \in \mathbf{F}$ is the unambiguous concatenation of two languages in $SD(\mathcal{C})$. Moreover, given $F \in \mathbf{F}$, we have $F = UH$ for $U \in \mathbf{U}$ which means that $\gamma(F) = \gamma(U)\gamma(H) \in T$ since T is closed under multiplication. It remains to show that our induction parameters have decreased. Since $\mathbf{F} = \{UH \mid U \in \mathbf{U}\}$ and $\gamma(U) \in Q_{\mathbf{H}}^*$ for every $U \in \mathbf{U}$ (by definition of \mathbf{U} in Lemma 44), one may verify that $Q_{\mathbf{F}}^+ \subseteq Q_{\mathbf{H}}^* \cdot \gamma(H)$. Hence, since $Q_{\mathbf{H}}^* \cdot \gamma(H) \subsetneq Q_{\mathbf{H}}^+$ by hypothesis in Sub-case 2, we have $Q_{\mathbf{F}}^+ \subsetneq Q_{\mathbf{H}}^+$. Our first induction parameter has decreased. Altogether, it follows that we may apply Proposition 40 in the case when P, \mathbf{H} and $t \in T$ have been replaced by L, \mathbf{F} and $1_Q \in T$. This yields a $SD(\mathcal{C})$ -partition \mathbf{V} of $L^* = ((P \setminus H)^*H)^*$ such that for every $V \in \mathbf{V}$, $V \in Q_{\mathbf{F}}^*$ and $\gamma(V) \in T$. Finally, it is clear by definition that $Q_{\mathbf{F}}^* \subseteq Q_{\mathbf{H}}^*$. Hence, the lemma follows. \blacktriangleleft

We are ready to construct the $SD(\mathcal{C})$ -partition \mathbf{K} of P^* and conclude the main argument. We let $\mathbf{K} = \{VU \mid V \in \mathbf{V} \text{ and } U \in \mathbf{U}\}$. It is immediate by definition that \mathbf{K} is a partition of P^* since P is a prefix code and \mathbf{V}, \mathbf{U} are partitions of $((P \setminus H)^*H)^*$ and $(P \setminus H)^*$ respectively (see the above discussion). Additionally, it is immediate by definition that \mathbf{K} is actually a $SD(\mathcal{C})$ -partition of P^* (it only contains unambiguous concatenations of languages in $SD(\mathcal{C})$). It remains to prove that \mathbf{K} satisfies (5): for every $K \in \mathbf{K}$, we have $\gamma(K) \in Q_{\mathbf{H}}^*$ and $t\gamma(K) \in T$. We have $K = VU$ with $V \in \mathbf{V}$ and $U \in \mathbf{U}$. By definition of \mathbf{U} and \mathbf{V} , we have $\gamma(U), \gamma(V) \in Q_{\mathbf{H}}^*$ and $\gamma(U), \gamma(V) \in T$. Moreover, $t \in T$ by hypothesis. Therefore, since both $Q_{\mathbf{H}}^*$ and T are closed under multiplication, it follows that $\gamma(K) \in Q_{\mathbf{H}}^*$ and $t\gamma(K) \in T$. This concludes the proof of Proposition 40.

C Appendix to Section 5

This appendix presents the missing proofs in Section 5. Similarly to what we did in Appendix A, we actually present a self-contained, full version of Section 5 which can be read independently. Recall that the purpose of Section 5 is to prove Theorem 3: when \mathcal{C} is a quotient-closed Boolean algebra of group languages with decidable separation, $SF(\mathcal{C})$ -covering is decidable.

As in Section 4, we rely on Proposition 19: we present an algorithm which computes $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from an input nice multiplicative rating map ρ . In this case as well, we do not work with $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ itself. Instead, we consider another set carrying more information. Defining this second object involves introducing a few additional concepts. We first present them and then turn to the algorithm. For details, see [26].

C.1 Preliminary definitions

Optimal ε -approximations. In this case, handling $SF(\mathcal{C})$ involves considering \mathcal{C} -optimal covers of $\{\varepsilon\}$. Since $\{\varepsilon\}$ is a singleton, there always exists such a cover consisting of a single language, which leads to the following definition.

Let \mathcal{C} be a Boolean algebra (we shall use the case when \mathcal{C} contains only group languages but this is not required for the definitions) and $\tau : 2^{A^*} \rightarrow Q$ be a rating map. A \mathcal{C} -optimal ε -approximation for τ is a language $L \in \mathcal{C}$ such that $\varepsilon \in L$ and $\tau(L) \leq \tau(L')$ for every $L' \in \mathcal{C}$ satisfying $\varepsilon \in L'$. As expected, there always exists a \mathcal{C} -optimal ε -approximation for any rating map τ .

► **Lemma 47.** *Let $\tau : 2^{A^*} \rightarrow Q$ be a rating map and \mathcal{C} be a Boolean algebra. There exists a \mathcal{C} -optimal ε -approximation for τ .*

Proof. Let $U \subseteq Q$ be the set of all elements $q \in Q$ such that $q = \tau(K)$ for some $K \in \mathcal{C}$ containing ε . Clearly, $\tau(A^*) \in U$ which means that U is non-empty since $A^* \in \mathcal{C}$ (\mathcal{C} is a Boolean algebra). For every $q \in U$, we fix an arbitrary language $K_q \in \mathcal{C}$ such that $\varepsilon \in K_q$ and $q = \tau(K_q)$ (K_q exists by definition of U). Finally, we let,

$$K = \bigcap_{q \in U} K_q$$

Since \mathcal{C} is a Boolean algebra, we have $K \in \mathcal{C}$. Moreover, $\varepsilon \in K$ by definition. Since $K \subseteq K_q$ for all $q \in U$, it follows that $\tau(K) \leq q$ for every $q \in U$. By definition of U , this implies that $\tau(K) \leq \tau(K')$ for every $K' \in \mathcal{C}$ such that $\varepsilon \in K'$. Hence, K is a \mathcal{C} -optimal ε -approximation for τ . ◀

Moreover, all \mathcal{C} -optimal ε -approximations for τ have the same image under τ . We write it $\mathfrak{i}_{\mathcal{C}}[\tau] \in Q$: $\mathfrak{i}_{\mathcal{C}}[\tau] = \tau(L)$ for every \mathcal{C} -optimal ε -approximation L for τ . It turns out that when τ is nice and multiplicative, computing $\mathfrak{i}_{\mathcal{C}}[\tau]$ from τ boils down to \mathcal{C} -separation. This is important: this is exactly how our algorithm for $SF(\mathcal{C})$ -covering depends on \mathcal{C} -separation.

► **Lemma 21.** *Let $\tau : 2^{A^*} \rightarrow Q$ be a nice rating map and \mathcal{C} a Boolean algebra. Then, $\mathfrak{i}_{\mathcal{C}}[\tau]$ is the sum of all $q \in Q$ such that $\{\varepsilon\}$ is not \mathcal{C} -separable from $\tau_*^{-1}(q)$.*

Proof. We let $U \subseteq Q$ be the set of all $q \in Q$ such that $\{\varepsilon\}$ is not \mathcal{C} -separable from $\tau_*^{-1}(q)$. Moreover, we let $r = \sum_{q \in U} q$. We show that $r = \mathfrak{i}_{\mathcal{C}}[\tau]$. First, we prove that $r \leq \mathfrak{i}_{\mathcal{C}}[\tau]$. By definition, this amounts to proving that given $q \in U$, we have $q \leq \mathfrak{i}_{\mathcal{C}}[\tau]$. By definition, $\mathfrak{i}_{\mathcal{C}}[\tau] = \tau(K)$ for some $K \in \mathcal{C}$ such that $\varepsilon \in K$. By definition of U , $\{\varepsilon\}$ is not \mathcal{C} -separable from $\tau_*^{-1}(q)$ which implies that $K \cap \tau_*^{-1}(q) \neq \emptyset$. Let $w \in K \cap \tau_*^{-1}(q)$. By definition $w \in K$ and $\tau(w) = q$. This implies that $q \leq \tau(K) = \mathfrak{i}_{\mathcal{C}}[\tau]$.

Conversely, we show that $\mathfrak{i}_{\mathcal{C}}[\tau] \leq r$. By definition, for every $q \in Q \setminus U$, $\{\varepsilon\}$ is \mathcal{C} -separable from $\tau_*^{-1}(q)$. We fix a language $H_q \in \mathcal{C}$ as a separator. We now let,

$$H = \bigcap_{q \in Q \setminus U} H_q \in \mathcal{C}$$

By definition, H separates $\{\varepsilon\}$ from $\tau_*^{-1}(q)$ for every $q \in Q \setminus U$. In particular, $\varepsilon \in H$ which implies that $\mathfrak{i}_{\mathcal{C}}[\tau] \leq \tau(H)$. Moreover, since τ is nice, we have $w_1, \dots, w_n \in H$ such that $\tau(H) = \tau(w_1) + \dots + \tau(w_n)$. Finally, since $H \cap \tau_*^{-1}(q) = \emptyset$ for every $q \in Q \setminus U$, we know that $\tau(w_1), \dots, \tau(w_n) \in U$. This implies that $\tau(H) \leq r$. Altogether, we get that $\mathfrak{i}_{\mathcal{C}}[\tau] \leq r$. ◀

Nested rating maps. We want an algorithm which computes $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from an input nice multiplicative rating map ρ for a fixed quotient-closed Boolean algebra of group languages \mathcal{C} . Yet, we shall not use optimal ε -approximations with this input rating map ρ . Instead, we consider an auxiliary rating map built from ρ (the definition is taken from [24]).

Consider a Boolean algebra \mathcal{D} (we shall use the case $\mathcal{D} = SF(\mathcal{C})$) and a rating map $\rho : 2^{A^*} \rightarrow R$. We build a new map $\xi_{\mathcal{D}}[\rho] : 2^{A^*} \rightarrow 2^R$ whose rating set is $(2^R, \cup)$. For every $K \subseteq A^*$, we define $\xi_{\mathcal{D}}[\rho](K) = \mathcal{I}_{\mathcal{D}}[K, \rho] \in 2^R$. It follows from Fact 18 that this is indeed a rating map (on the other hand $\xi_{\mathcal{D}}[\rho]$ need not be nice, see [24] for a counterexample). More importantly, $\xi_{\mathcal{D}}[\rho]$ need not be multiplicative, even if this is the case for $\rho : 2^{A^*} \rightarrow R$. However, it turns out that we are able to cope with this last issue.

Clearly, when $\rho : 2^{A^*} \rightarrow R$ is a multiplicative rating map, the set 2^R is an idempotent semiring (addition is union and the multiplication is lifted from the one of R). It turns out that when \mathcal{D} is quotient-closed Boolean algebra closed under concatenation (such as when $\mathcal{D} = SF(\mathcal{C})$), $\xi_{\mathcal{D}}[\rho] : 2^{A^*} \rightarrow 2^R$ behaves almost as a multiplicative rating map for this semiring structure on 2^R . Indeed, we have the following statement taken from [24] (see Lemma 6.7). For $S \subseteq R$, we write $\downarrow_R S \subseteq R$ for the set $\downarrow_R S = \{q \mid q \leq r \text{ for some } r \in S\}$.

► **Lemma 48.** *Consider a quotient-closed Boolean algebra \mathcal{C} and a multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. Then, for every $H, K \subseteq A^*$, we have $\downarrow_R \xi_{SF(\mathcal{C})}[\rho](K) = \xi_{SF(\mathcal{C})}[\rho](K)$ and $\downarrow_R(\xi_{SF(\mathcal{C})}[\rho](H) \cdot \xi_{SF(\mathcal{C})}[\rho](K)) = \xi_{SF(\mathcal{C})}[\rho](HK)$.*

We may now explain which set is computed by our algorithm instead of $\mathcal{I}_{SF(\mathcal{C})}[\rho]$. Consider a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. Since $\xi_{SF(\mathcal{C})}[\rho] : 2^{A^*} \rightarrow 2^R$ is a rating map, we may consider the element $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \in 2^R$. By definition, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] = \xi_{SF(\mathcal{C})}[\rho](L)$ where L is a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$. Therefore, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is a subset of $\xi_{SF(\mathcal{C})}[\rho](A^*) = \mathcal{I}_{SF(\mathcal{C})}[A^*, \rho] = \mathcal{I}_{SF(\mathcal{C})}[\rho]$. When \mathcal{C} is a quotient-closed Boolean algebra of group languages, one may compute the whole set $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from this subset.

► **Proposition 22.** *Let \mathcal{C} be a quotient-closed Boolean algebra of group languages and $\rho : 2^{A^*} \rightarrow R$ a nice multiplicative rating map. Then, $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ is the least subset of R containing $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ and satisfying the three following properties:*

- **Trivial elements.** *For every $w \in A$, $\rho(w) \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$.*
- **Downset.** *For every $r \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$ and $q \leq r$, we have $q \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$.*
- **Multiplication.** *For every $q, r \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$, we have $qr \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$.*

Proof. We let $S \subseteq R$ be the least subset of R containing $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ and satisfying the three properties in the proposition. We need to prove that $S = \mathcal{I}_{SF(\mathcal{C})}[\rho]$. The inclusion $S \subseteq \mathcal{I}_{SF(\mathcal{C})}[\rho]$ is immediate. We already observed before the statement that $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \subseteq \mathcal{I}_{SF(\mathcal{C})}[\rho]$. The other properties are generic ones of optimal imprints when the investigated class is a quotient-closed Boolean algebra (see Lemma 9.5 in [22]).

We concentrate on the converse inclusion: $\mathcal{I}_{SF(\mathcal{C})}[\rho] \subseteq S$. For the proof, we let $L \subseteq A^*$ be a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$. Recall that by definition, this means that $L \in \mathcal{C}$, $\varepsilon \in L$ and $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] = \xi_{SF(\mathcal{C})}[\rho](L)$. Consider $r \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$, we show that $r \in S$. The argument is based on the following lemma which is where we use the hypothesis that \mathcal{C} is made of *group languages*.

► **Lemma 49.** *There exist $\ell \in \mathbb{N}$ and $a_1, \dots, a_\ell \in A$ such that $r \in \xi_{SF(\mathcal{C})}[\rho](La_1L \cdots a_\ell L)$.*

Proof. Since $L \in \mathcal{C}$, it is a group language by hypothesis on \mathcal{C} : it is recognized by a finite group. Therefore, since $\varepsilon \in L$, one may use a pumping argument to show that A^* is a finite union of languages having the form $La_1L \cdots a_\ell L$ for $a_1, \dots, a_\ell \in A$. Moreover, $\mathcal{I}_{SF(\mathcal{C})}[\rho] = \mathcal{I}_{SF(\mathcal{C})}[A^*, \rho] = \xi_{SF(\mathcal{C})}[\rho](A^*)$ by definition. Therefore, since $\xi_{SF(\mathcal{C})}[\rho]$ is a rating map (whose rating set is $(2^R, \cup)$), it follows that $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ is a finite union of sets $\xi_{SF(\mathcal{C})}[\rho](La_1L \cdots a_\ell L)$ for $a_1, \dots, a_\ell \in A$. Since $r \in \mathcal{I}_{SF(\mathcal{C})}[\rho]$, the result follows. ◀

In view of Lemma 48, the hypothesis that $r \in \xi_{SF(\mathcal{C})}[\rho](La_1L \cdots a_\ell L)$ given by Lemma 49 implies that,

$$r \in \downarrow_R(\xi_{SF(\mathcal{C})}[\rho](L) \cdot \xi_{SF(\mathcal{C})}[\rho](a_1) \cdot \xi_{SF(\mathcal{C})}[\rho](L) \cdots \xi_{SF(\mathcal{C})}[\rho](a_\ell) \cdot \xi_{SF(\mathcal{C})}[\rho](L))$$

By definition of L , we have $\xi_{SF(\mathcal{C})}[\rho](L) = \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. Moreover, we have the following fact.

► **Fact 50.** For every $a \in A$, we have $\xi_{SF(\mathcal{C})}[\rho](a) = \downarrow_R\{\rho(a)\}$.

Proof. By definition, $\xi_{SF(\mathcal{C})}[\rho](a) = \mathcal{I}_{SF(\mathcal{C})}[\{a\}, \rho]$. Since $\{a\} \in SF(\mathcal{C})$ by definition, it is clear that $\{\{a\}\}$ is an optimal $SF(\mathcal{C})$ -cover of $\{a\}$. Thus, $\mathcal{I}_{SF(\mathcal{C})}[\{a\}, \rho] = \mathcal{I}[\rho](\{a\}) = \downarrow_R\{\rho(a)\}$ and we are finished. ◀

Altogether, we obtain that,

$$r \in \downarrow_R(\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \cdot \{\rho(a_1)\} \cdot \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \cdots \{\rho(a_\ell)\} \cdot \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]])$$

It is now immediate from the definition of S that $r \in S$ which concludes the proof for the inclusion $\mathcal{I}_{SF(\mathcal{C})}[\rho] \subseteq S$. ◀

C.2 Algorithm

We may now present our algorithm for computing $\mathcal{I}_{SF(\mathcal{C})}[\rho]$. We fix a quotient-closed Boolean algebra of group languages \mathcal{C} for the presentation. As expected, the main procedure computes $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ (see Proposition 22). In this case as well, this procedure is obtained from a characterization theorem.

Consider a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. We define the $SF(\mathcal{C})$ -complete subsets of R for ρ . The definition depends on auxiliary nice multiplicative rating maps. We first present them. Clearly, 2^R is an idempotent semiring (addition is union and the multiplication is lifted from the one of R). For every $S \subseteq R$, we use it as the rating set of a nice multiplicative rating map $\eta_{\rho, S} : 2^{A^*} \rightarrow 2^R$. Since we are defining a *nice* multiplicative rating map, it suffices to specify the evaluation of letters. For $a \in A$, we let $\eta_{\rho, S}(a) = S \cdot \{\rho(a)\} \cdot S \in 2^R$. Observe that by definition, we have $\mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}] \subseteq R$.

We are ready to define the $SF(\mathcal{C})$ -complete subsets of R . Consider $S \subseteq R$. We say that S is $SF(\mathcal{C})$ -complete for ρ when the following conditions are satisfied:

1. **Downset.** For every $r \in S$ and $q \leq r$, we have $q \in S$.
2. **Multiplication.** For every $q, r \in S$, we have $qr \in S$.
3. **\mathcal{C} -operation.** We have $\mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}] \subseteq S$.
4. **$SF(\mathcal{C})$ -closure.** For every $r \in S$, we have $r^\omega + r^{\omega+1} \in S$.

► **Remark 24.** The definition of $SF(\mathcal{C})$ -complete subsets does not explicitly require that they contain some trivial elements. Yet, this is implied by \mathcal{C} -operation. Indeed, if $S \subseteq R$ is $SF(\mathcal{C})$ -complete, then $\eta_{\rho, S}(\varepsilon) = \{1_R\}$ (this is the multiplicative neutral element of 2^R). This implies that $1_R \in \mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}]$ and we obtain from \mathcal{C} -operation that $1_R \in S$.

► **Theorem 25** (*$SF(\mathcal{C})$ -optimal imprints (\mathcal{C} made of group languages)*). Let $\rho : 2^{A^*} \rightarrow R$ be a nice multiplicative rating map. Then, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is the least $SF(\mathcal{C})$ -complete subset of R .

When \mathcal{C} -separation is decidable, Theorem 25 yields a least fixpoint procedure for computing $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ from a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$. The computation starts from the empty set and saturates it with the four operations in the definition of $SF(\mathcal{C})$ -complete subsets. It is clear that we may implement downset, multiplication and $SF(\mathcal{C})$ -closure. Moreover, we may implement \mathcal{C} -operation as this boils down to \mathcal{C} -separation by Lemma 21. Eventually, the computation reaches a fixpoint and it is straightforward to verify that this set is the least $SF(\mathcal{C})$ -complete subset of R , *i.e.*, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ by Theorem 25.

By Proposition 22, we may compute $\mathcal{I}_{SF(\mathcal{C})}[\rho]$ from $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. Altogether, this yields the decidability of $SF(\mathcal{C})$ -covering by Proposition 19. Hence, Theorem 3 is proved.

C.3 Proof

We now concentrate on proving Theorem 25. We fix a nice multiplicative rating map $\rho : 2^{A^*} \rightarrow R$ for the proof. We need to show that $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is the least $SF(\mathcal{C})$ -complete subset of R . The proof involves soundness and completeness directions. In both cases, we apply Theorem 20 as a sub-result.

Soundness

We prove that $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is $SF(\mathcal{C})$ -complete. We start with a preliminary simple fact that will be useful.

► **Fact 51.** *There exists a finite group G and a morphism $\alpha : A^* \rightarrow G$ such that the language $L = \alpha^{-1}(1_G)$ is a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$.*

Proof. We let H be a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$: we have $H \in \mathcal{C}$, $\varepsilon \in H$ and $\xi_{SF(\mathcal{C})}[\rho](H) = \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. Let $\alpha : A^* \rightarrow G$ be the syntactic morphism of H . Since \mathcal{C} is a quotient-closed Boolean algebra of group languages and $H \in \mathcal{C}$, it is standard that G is a finite group and that every language recognized by α belongs to \mathcal{C} (see [15] for example). In particular $L = \alpha^{-1}(1_G) \in \mathcal{C}$. Moreover, since $\varepsilon \in H$ and H is recognized by α , we have $L \subseteq H$. Since H is a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$, L must be one as well. ◀

We fix the morphism $\alpha : A^* \rightarrow G$ and the \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$ $L = \alpha^{-1}(1_G)$ for the proof. We need to show that $\xi_{SF(\mathcal{C})}[\rho](L) = \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is $SF(\mathcal{C})$ -complete. There are four properties to verify.

Downset (Condition 1). By definition of L , we have $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] = \xi_{SF(\mathcal{C})}[\rho](L)$ and it is immediate from Lemma 48 that $\xi_{SF(\mathcal{C})}[\rho](L) = \downarrow_R \xi_{SF(\mathcal{C})}[\rho](L)$.

Multiplication (Condition 2). Let $q, r \in \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. We show that $qr \in \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. By hypothesis, we have $q, r \in \xi_{SF(\mathcal{C})}[\rho](L)$. In particular, this implies that,

$$qr \in \downarrow_R(\xi_{SF(\mathcal{C})}[\rho](L) \cdot \xi_{SF(\mathcal{C})}[\rho](L))$$

Hence, Lemma 48 yields that $qr \in \xi_{SF(\mathcal{C})}[\rho](LL)$. Finally, since $L = \alpha^{-1}(1_G)$, we have $LL = L$. Thus, $qr \in \xi_{SF(\mathcal{C})}[\rho](L) = \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$.

\mathcal{C} -operation (Condition 3). For the sake of avoiding clutter, we write $S = \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. We need to show that $\mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}] \subseteq S$. Hence, we consider $r \in \mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}]$ and show that $r \in S = \xi_{SF(\mathcal{C})}[\rho](L)$. Since $L \in \mathcal{C}$ and $\varepsilon \in L$, the hypothesis that $r \in \mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}]$ yields $r \in \eta_{\rho, S}(L)$. Hence, we get $w \in L$ such that $r \in \eta_{\rho, S}(w)$. There are now two cases depending on whether $w = \varepsilon$ or $w \in A^+$.

Assume first that $w = \varepsilon$. In that case, we have $\eta_{\rho, S}(w) = \{1_R\}$ (this is the multiplicative neutral element of 2^R). Thus, $r = 1_R$. Since $\varepsilon \in L$, it is clear that $1_R \in \mathcal{I}_{SF(\mathcal{C})}[L, \rho] = \xi_{SF(\mathcal{C})}[\rho](L) = S$ which concludes this case.

We now assume that $w \in A^+$. In that case, there exist $n \geq 1$ and $a_1, \dots, a_n \in A$ such that $w = a_1 \cdots a_n$. By definition $\eta_{\rho, S}(a) = S \cdot \{\rho(a_1)\} \cdot S$ for every $a \in A$. Moreover, since we already established that $S = \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is closed under multiplication, we have $S \cdot S \subseteq S$. Hence, it follows that

$$\eta_{\rho, S}(w) \subseteq S \cdot \{\rho(a_1)\} \cdot S \cdots \{\rho(a_n)\} \cdot S$$

We get $r \in S \cdot \{\rho(a_1)\} \cdot S \cdots \{\rho(a_n)\} \cdot S$. By definition, we have $S = \xi_{SF(\mathcal{C})}[\rho](L)$ and it is clear that for every $a \in A$, we have $\rho(a) \in \mathcal{I}_{SF(\mathcal{C})}[\{a\}, \rho] = \xi_{SF(\mathcal{C})}[\rho](a)$. Thus, we obtain that,

$$r \in \xi_{SF(\mathcal{C})}[\rho](L) \cdot \prod_{1 \leq i \leq n} (\xi_{SF(\mathcal{C})}[\rho](a_i) \cdot \xi_{SF(\mathcal{C})}[\rho](L))$$

By Lemma 48, we have,

$$\xi_{SF(\mathcal{C})}[\rho](La_1L \cdots La_nL) = \downarrow_R \left(\xi_{SF(\mathcal{C})}[\rho](L) \cdot \prod_{1 \leq i \leq n} (\xi_{SF(\mathcal{C})}[\rho](a_i) \cdot \xi_{SF(\mathcal{C})}[\rho](L)) \right)$$

Consequently, $r \in \xi_{SF(\mathcal{C})}[\rho](La_1L \cdots La_nL)$. Recall that $L = \alpha^{-1}(1_G)$ and $a_1 \cdots a_n \in L$ (i.e., $\alpha(a_1 \cdots a_n) = 1_G$). Thus, α maps every word in $La_1L \cdots La_nL$ to 1_G and we obtain that $La_1L \cdots La_nL \subseteq L$. Since $\xi_{SF(\mathcal{C})}[\rho]$ is a rating map, this yields $\xi_{SF(\mathcal{C})}[\rho](La_1L \cdots La_nL) \subseteq \xi_{SF(\mathcal{C})}[\rho](L)$. Altogether, we obtain $r \in \xi_{SF(\mathcal{C})}[\rho](L) = S$, finishing the proof.

$SF(\mathcal{C})$ -closure (Condition 4). Let $r \in \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$. We show that $r^\omega + r^{\omega+1} \in \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ as well. Since L is a \mathcal{C} -optimal ε -approximation for $\xi_{SF(\mathcal{C})}[\rho]$, we need to show that:

$$r^\omega + r^{\omega+1} \in \xi_{SF(\mathcal{C})}[\rho](L) = \mathcal{I}_{SF(\mathcal{C})}[L, \rho].$$

Let \mathbf{K} be an optimal $SF(\mathcal{C})$ -cover of L for ρ . We need to exhibit $K \in \mathbf{K}$ such that $r^\omega + r^{\omega+1} \leq \rho(K)$. This is where we use Theorem 20. First, we use \mathbf{K} and L to construct a finite sub-class of \mathcal{C} .

► **Fact 52.** *There exists a finite quotient-closed Boolean algebra $\mathcal{G} \subseteq \mathcal{C}$ such that $L \in \mathcal{G}$ and $K \in SF(\mathcal{G})$ for every $K \in \mathbf{K}$.*

Proof. By definition \mathbf{K} contains finitely many languages in $SF(\mathcal{C})$. Hence, there exists a finite set of languages $\mathbf{H} \subseteq \mathcal{C}$ such that every language $K \in \mathbf{K}$ is built by applying Boolean operations and concatenations to languages in \mathbf{H} . Since \mathcal{C} is a quotient-closed Boolean algebra, it is standard that there exists a finite quotient-closed Boolean algebra $\mathcal{G} \subseteq \mathcal{C}$ which contains all languages in \mathbf{H} and L (since $L \in \mathcal{C}$ by definition). See Lemma 17 in [23] for a proof. It is then immediate by definition of \mathbf{H} that $K \in SF(\mathcal{G})$ for every $K \in \mathbf{K}$. ◀

Since \mathcal{G} is finite, we may consider the associated equivalence $\sim_{\mathcal{G}}$ defined on A^* . By definition, we have $[\varepsilon]_{\mathcal{G}} \in \mathcal{G} \subseteq \mathcal{C}$ and $\varepsilon \in [\varepsilon]_{\mathcal{G}}$. Thus, $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \subseteq \xi_{SF(\mathcal{C})}[\rho]([\varepsilon]_{\mathcal{G}})$. Since $r \in \mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ by hypothesis, this yields,

$$r \in \xi_{SF(\mathcal{C})}[\rho]([\varepsilon]_{\mathcal{G}}) = \mathcal{I}_{SF(\mathcal{C})}[[\varepsilon]_{\mathcal{G}}, \rho].$$

Moreover, we have $\mathcal{G} \subseteq \mathcal{C}$ which implies that $SF(\mathcal{G}) \subseteq SF(\mathcal{C})$. Consequently, Fact 18 yields that $r \in \mathcal{I}_{SF(\mathcal{G})}[[\varepsilon]_{\mathcal{G}}, \rho]$. By definition, this property can be reformulated as follows,

$$([\varepsilon]_{\mathcal{G}}, r) \in \mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}}[\rho]$$

By Theorem 20, we know that the set $\mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}}[\rho]$ is $SF(\mathcal{G})$ -saturated for ρ . Hence, it satisfies $SF(\mathcal{G})$ -closure and since $[\varepsilon]_{\mathcal{G}}$ is an idempotent $\sim_{\mathcal{G}}$ -class, we get that:

$$([\varepsilon]_{\mathcal{G}}, r^\omega + r^{\omega+1}) \in \mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}}[\rho].$$

Therefore, we have $r^\omega + r^{\omega+1} \in \mathcal{I}_{SF(\mathcal{G})}[[\varepsilon]_{\mathcal{G}}, \rho]$.

121:38 On all things star-free

Finally, we know that $L \in \mathcal{G}$ by definition of \mathcal{G} (see Fact 52). Thus, the hypothesis that $\varepsilon \in L$ yields $[\varepsilon]_{\mathcal{G}} \subseteq L$. We then obtain from Fact 18 that $\mathcal{I}_{SF(\mathcal{G})}([\varepsilon]_{\mathcal{G}}, \rho) \subseteq \mathcal{I}_{SF(\mathcal{G})}[L, \rho]$. By definition \mathbf{K} is a cover of L . Moreover, it is an $SF(\mathcal{G})$ -cover by definition of \mathcal{G} in Fact 52. Thus, $\mathcal{I}_{SF(\mathcal{G})}[L, \rho] \subseteq \mathcal{I}[\rho](\mathbf{K})$. Altogether, we obtain that $r^\omega + r^{\omega+1} \in \mathcal{I}[\rho](\mathbf{K})$ which yields $K \in \mathbf{K}$ such that $r^\omega + r^{\omega+1} \leq \rho(K)$, finishing the soundness proof.

Completeness

We have proved that $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]]$ is a $SF(\mathcal{C})$ -complete subset of R . It remains to show that it is the least such subset. Hence, we fix an arbitrary $SF(\mathcal{C})$ -complete set $S \subseteq R$ and show that $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \subseteq S$. By definition, we know that $\mathfrak{i}_{\mathcal{C}}[\xi_{SF(\mathcal{C})}[\rho]] \subseteq \xi_{SF(\mathcal{C})}[\rho](L)$ for every language $L \in \mathcal{C}$ such that $\varepsilon \in L$. Hence, it suffices to exhibit a language $L \in \mathcal{C}$ such that $\varepsilon \in L$ and $\xi_{SF(\mathcal{C})}[\rho](L) \subseteq S$. We first choose the appropriate language L .

We let H be a \mathcal{C} -optimal ε -approximation for the nice multiplicative rating map $\eta_{\rho, S}$. That is, we have $H \in \mathcal{C}$, $\varepsilon \in H$ and $\eta_{\rho, S}(H) = \mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}]$. Since $H \in \mathcal{C}$, one may verify that there exists a **finite** quotient-closed Boolean algebra $\mathcal{G} \subseteq \mathcal{C}$ such that $H \in \mathcal{G}$ (see again Lemma 17 in [23] for a proof). We choose $L = [\varepsilon]_{\mathcal{G}}$: clearly, this language belongs to $\mathcal{G} \subseteq \mathcal{C}$ and contains ε by definition. It now remains to show the following inclusion:

$$\xi_{SF(\mathcal{C})}[\rho]([\varepsilon]_{\mathcal{G}}) \subseteq S \quad (9)$$

Let us give a brief overview of the proof. It is based on Theorem 20. First, we use our set $S \subseteq R$ which is $SF(\mathcal{C})$ -complete subset (for ρ) to build another set $S' \subseteq (A^*/\sim_{\mathcal{G}}) \times R$ which is $SF(\mathcal{G})$ -saturated (for ρ). This is where we apply Theorem 20. It states that the *least* $SF(\mathcal{G})$ -saturated set is $\mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}}[\rho]$. Hence, we obtain that the inclusion $\mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}}[\rho] \subseteq S'$ holds. It is then straightforward to prove (9) from this inclusion.

Let us first define the set $S' \subseteq (A^*/\sim_{\mathcal{G}}) \times R$. The construction is based on the set S and the nice multiplicative rating map $\eta_{\rho, S}$. We define,

$$S' = \{([\varepsilon]_{\mathcal{G}}, s) \mid s \in S\} \cup \{(C, r) \mid C \in A^*/\sim_{\mathcal{G}} \text{ and } r \in \downarrow_R \eta_{\rho, S}(C)\}$$

Observe that by definition, we have the following useful fact about this new set S' .

► **Fact 53.** *For every $s \in R$, if $([\varepsilon]_{\mathcal{G}}, s) \in S'$, then $s \in S$.*

Proof. Let $s \in R$ such that $([\varepsilon]_{\mathcal{G}}, s) \in S'$. By definition of S' , either $s \in S$ or $s \in \downarrow_R \eta_{\rho, S}([\varepsilon]_{\mathcal{G}})$. In the former case, we are finished. Hence, we assume that $s \in \downarrow_R \eta_{\rho, S}([\varepsilon]_{\mathcal{G}})$. By definition, $[\varepsilon]_{\mathcal{G}} \subseteq H$ (we have $H \in \mathcal{G}$ and $\varepsilon \in H$). Since H is a \mathcal{C} -optimal ε -approximation for $\eta_{\rho, S}$, this yields $\eta_{\rho, S}([\varepsilon]_{\mathcal{G}}) \subseteq \mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}]$. hence, we have $s \in \downarrow_R \mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}]$. Finally, since S is $SF(\mathcal{C})$ -complete, we know that $\mathfrak{i}_{\mathcal{C}}[\eta_{\rho, S}] \subseteq S$ and $\downarrow_R S = S$. Altogether, we get that $s \in S$, concluding the proof. ◀

We now turn to the technical core of the proof: our new set $S' \subseteq (A^*/\sim_{\mathcal{G}}) \times R$ is $SF(\mathcal{G})$ -saturated for ρ . We state this result in the following lemma.

► **Lemma 54.** *The set $S' \subseteq (A^*/\sim_{\mathcal{G}}) \times R$ is $SF(\mathcal{G})$ -saturated for ρ .*

Before we prove Lemma 54, let us use it to show that (9) holds and conclude the completeness proof. We need to show that $\xi_{SF(\mathcal{C})}[\rho]([\varepsilon]_{\mathcal{G}}) \subseteq S$. Hence, we consider $r \in \xi_{SF(\mathcal{C})}[\rho]([\varepsilon]_{\mathcal{G}})$ and prove that $r \in S$.

By definition, we know that $r \in \mathcal{I}_{SF(\mathcal{C})} [[\varepsilon]_{\mathcal{G}}, \rho]$. Moreover, since $\mathcal{G} \subseteq \mathcal{C}$, we have $SF(\mathcal{G}) \subseteq SF(\mathcal{C})$. This yields the following inclusion,

$$\mathcal{I}_{SF(\mathcal{C})} [[\varepsilon]_{\mathcal{G}}, \rho] \subseteq \mathcal{I}_{SF(\mathcal{G})} [[\varepsilon]_{\mathcal{G}}, \rho]$$

Consequently, we get $r \in \mathcal{I}_{SF(\mathcal{G})} [[\varepsilon]_{\mathcal{G}}, \rho]$ which can be reformulated as follows:

$$([\varepsilon]_{\mathcal{G}}, r) \in \mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}} [\rho].$$

Theorem 20 states that $\mathcal{P}_{SF(\mathcal{G})}^{\mathcal{G}} [\rho]$ is the least $SF(\mathcal{G})$ -saturated subset of $(A^*/\sim_{\mathcal{G}}) \times R$. Hence, since S' is also $SF(\mathcal{G})$ -saturated by Lemma 54, we obtain that $([\varepsilon]_{\mathcal{G}}, r) \in S'$. By Fact 53, this implies that $r \in S$, concluding the main argument. It remains to prove Lemma 54.

Proof of Lemma 54. We show that $S' \subseteq (A^*/\sim_{\mathcal{G}}) \times R$ is $SF(\mathcal{G})$ -saturated for ρ . This involves four properties. They are proved independently.

Trivial elements. Consider $w \in A^*$, we have to show that $([w]_{\mathcal{G}}, \rho(w)) \in S'$. By definition of S' , it suffices to prove that $\rho(w) \in \downarrow_R \eta_{\rho, S}([w]_{\mathcal{G}})$. Clearly, $w \in [w]_{\mathcal{G}}$. Hence, it now suffices to prove that $\rho(w) \in \eta_{\rho, S}(w)$. If $w = \varepsilon$, then $\rho(w) = 1_R$ and $\eta_{\rho, S}(\varepsilon) = \{1_R\}$ (this is the neutral element of 2^R). Thus, it is immediate that $\rho(w) \in \eta_{\rho, S}(w)$. Assume now that $w \in A^+$. We have $a_1, \dots, a_n \in A$ such that $w = a_1 \cdots a_n$. Since S is $SF(\mathcal{C})$ -complete, we have $1_R \in S$ (this is obtained from \mathcal{C} -operation, see Remark 24). Hence, we get that for every $i \leq n$, $\rho(a_i) \in S \cdot \{\rho(a_i)\} \cdot S = \eta_{\rho, S}(a_i)$. It then follows that $\rho(w) \in \eta_{\rho, S}(w)$ which concludes this case.

Downset. Consider $(C, r) \in S'$ and $q \leq r$. We show that $(C, q) \in S'$. By definition, there are two possible cases. First, it may happen that $C = [\varepsilon]_{\mathcal{G}}$ and $r \in S$. In that case, $q \in S$ since $\downarrow_R S = S$ (S is $SF(\mathcal{C})$ -complete) and we get that $(C, q) = ([\varepsilon]_{\mathcal{G}}, q) \in S'$. Otherwise, $r \in \downarrow_R \eta_{\rho, S}(C)$ which means that $q \in \downarrow_R \eta_{\rho, S}(C)$ as well and we get $(C, q) \in S'$, concluding the proof for downset.

Multiplication. Consider $(C, q), (D, r) \in S'$. We have to show that $(C \cdot D, qr) \in S'$. By definition of S' , there are several cases.

Assume first that $C = D = [\varepsilon]_{\mathcal{G}}$. In that case, Fact 53 yields that $q, r \in S$. Since S is $SF(\mathcal{C})$ -complete, this implies that $qr \in S$ and we obtain that $(C \cdot D, qr) = ([\varepsilon]_{\mathcal{G}}, qr) \in S'$. We now assume that either C or D is distinct from $[\varepsilon]_{\mathcal{G}}$ for the remainder of this case.

Assume first that both C and D are distinct from $[\varepsilon]_{\mathcal{G}}$. By definition of S' , this implies that $q \in \downarrow_R \eta_{\rho, S}(C)$ and $r \in \downarrow_R \eta_{\rho, S}(D)$. It follows that $qr \in \downarrow_R \eta_{\rho, S}(CD)$. Since $CD \subseteq C \cdot D$, this yields $qr \in \downarrow_R \eta_{\rho, S}(C \cdot D)$ and we get $(C \cdot D, qr) \in S'$.

Finally, we handle the case when $C = [\varepsilon]_{\mathcal{G}}$ and $D \neq [\varepsilon]_{\mathcal{G}}$ (the symmetrical case is left to the reader). By Fact 53, the hypothesis that $C = [\varepsilon]_{\mathcal{G}}$ yields $q \in S$. Moreover, we have $r \in \downarrow_R \eta_{\rho, S}(D)$ since $D \neq [\varepsilon]_{\mathcal{G}}$. Observe that the hypothesis $D \neq [\varepsilon]_{\mathcal{G}}$ also implies that $D \subseteq A^+$. Since $S \cdot S \subseteq S$ (S is $SF(\mathcal{C})$ -complete), one may verify from the definition of $\eta_{\rho, S}$ that this yields $S \cdot \eta_{\rho, S}(D) \subseteq \eta_{\rho, S}(D)$. Thus, since $q \in S$ and $r \in \downarrow_R \eta_{\rho, S}(D)$, we get $qr \in \downarrow_R \eta_{\rho, S}(D)$. Finally, $C \cdot D = D$ since $C = [\varepsilon]_{\mathcal{G}}$. Thus, $qr \in \downarrow_R \eta_{\rho, S}(C \cdot D)$ which yields $(C \cdot D, qr) \in S'$, concluding the proof for multiplication.

$SF(\mathcal{G})$ -closure. It remains to handle $SF(\mathcal{G})$ -closure. Since \mathcal{G} is a sub-class of \mathcal{C} , it is a quotient-closed Boolean algebra of group languages. Thus, one may verify that $A^*/\sim_{\mathcal{G}}$ is a group which implies that $[\varepsilon]_{\mathcal{G}}$ is the only idempotent of $A^*/\sim_{\mathcal{G}}$. Hence, it suffices to show that for every $r \in R$ such that $([\varepsilon]_{\mathcal{G}}, r) \in S'$, we have $([\varepsilon]_{\mathcal{G}}, r^\omega + r^{\omega+1}) \in S'$. By Fact 53, we have $r \in S$. Thus, since S is $SF(\mathcal{C})$ -complete, $SF(\mathcal{C})$ -closure yields that $r^\omega + r^{\omega+1} \in S$ which implies $([\varepsilon]_{\mathcal{G}}, r^\omega + r^{\omega+1}) \in S'$ by definition, finishing the proof. \blacktriangleleft